A Matemática do Ensino Médio

Volume 1

Elon Lages Lima
Paulo Cezar Pinto Carvalho
Eduardo Wagner
Augusto César Morgado



COLEÇÃO DO PROFESSOR DE MATEMÁTICA
SOCIEDADE BRASILEIRA DE MATEMÁTICA

trada Dona Castorina, 110 - Horto
460-320, Rio de Janeiro - RJ

ıpresso no Brasil / Printed in Brazil

oleção do Professor de Matemática

ıpa: Rodolfo Capeto
ıstração: Tina Velho

stribuição e vendas:
ciedade Brasileira de Matemática
nail: vendalivros@sbm.org.br
l.: (21) 2529-5073
vw.sbm.org.br

3N: 85-8581810-7

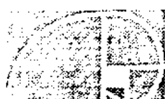
A Matemática do Ensino Médio

Volume 1

Oitava Edição

Elon Lages Lima
Paulo Cezar Pinto Carvalho
Eduardo Wagner
Augusto Cesar Morgado

Coleção do Professor de Matemática
Sociedade Brasileira de Matemática, Rio de Janeiro



COLEÇÃO DO PROFESSOR DE MATEMÁTICA

- *Logaritmos* - E.L.Lima
- *Análise Combinatória e Probabilidade com as soluções dos exercícios* - A.C.Morgado, J.B.Pitombeira, P.C.P.Carvalho e P.Fernandez
- *Medida e Forma em Geometria (Comprimento, Área, Volume e Semelhança)* - E.L.Lima
- *Meu Professor de Matemática e outras Histórias* - E.L.Lima
- *Coordenadas no Plano com as soluções dos exercícios* - E.L.Lima com a colaboração de P.C.P.Carvalho
- *Trigonometria, Números Complexos* - M.P.do Carmo, A.C.Morgado, E.Wagner, Notas Históricas de J.B.Pitombeira
- *Coordenadas no Espaço* - E.L.Lima
- *Progressões e Matemática Financeira* - A.C.Morgado, E.Wagner e S.C.Zani
- *Construções Geométricas* - E.Wagner com a colaboração de J.P.Q.Carneiro
- *Introdução à Geometria Espacial* - P.C.P.Carvalho
- *Geometria Euclidiana Plana* - J.L.M.Barbosa
- *Isometrias* - E.L.Lima
- *A Matemática do Ensino Médio Vol. 1* - E.L.Lima, P.C.P.Carvalho, E.Wagner e A.C.Morgado
- *A Matemática do Ensino Médio Vol. 2* - E.L.Lima, P.C.P.Carvalho, E.Wagner e A.C.Morgado
- *A Matemática do Ensino Médio Vol. 3* - E.L.Lima, P.C.P.Carvalho, E.Wagner e A.C.Morgado
- *Matemática e Ensino* - E.L.Lima
- *Temas e Problemas* - E.L.Lima, P.C.P.Carvalho, E.Wagner e A.C.Morgado
- *Episódios da História Antiga da Matemática* - A.Aaboe
- *Exame de Textos: Análise de livros de Matemática* - E.L.Lima

COLEÇÃO INICIAÇÃO CIENTÍFICA

- *Números Irracionais e Transcendentes* - D.G.de Figueiredo
- *Primalidade em Tempo Polinomial - Uma Introdução ao Algoritmo AKS* - S.C.Coutinho

COLEÇÃO TEXTOS UNIVERSITÁRIOS

- *Introdução à Computação Algébrica com o Maple* - L.N.de Andrade
- *Elementos de Aritmética* - A. Hefez
- *Métodos Matemáticos para a Engenharia* - E.C.de Oliveira e M.Tygel

COLEÇÃO MATEMÁTICA APLICADA

- *Introdução à Inferência Estatística* - H.Bolfarine e M.Sandoval

COLEÇÃO OLIMPÍADAS

- *Olimpíadas Brasileiras de Matemática*, 9ª a 16ª - C.Moreira, E.Motta, E.Tengan, L.Amâncio, N.Saldanha, P.Rodrigues

Conteúdo

Capítulo 6 - Funções Quadráticas

Capítulo 7 - Funções Polinomiais

Capítulo 8 - Funções Exponenciais e Logarítmicas

Capítulo 9 - Funções Trigonométricas

Prefácio

O programa de Matemática da primeira série do Ensino Médio tem como tema central as funções reais de uma variável real, estudadas sob o ponto de vista elementar, isto é, sem o uso do Cálculo Infinitesimal. Como preliminar a esse estudo e preparação para as séries subseqüentes, são apresentadas noções sobre conjuntos, a idéia geral de função e as diferentes categorias de números (naturais, inteiros, racionais e, principalmente, reais).

O presente livro cobre esse programa. Ele contém a matéria lecionada no primeiro dos três módulos do curso de aperfeiçoamento para professores de Matemática, iniciado no segundo semestre de 1996, no IMPA, tendo como instrutores os professores A.C.O. Morgado, E. Wagner, Paulo Cézar Carvalho e o autor. A estes caros amigos e competentes colaboradores devo uma revisão crítica do manuscrito, a sugestão de alguns exemplos interessantes e a inclusão de numerosos exercícios. Por essa valiosa participação, registro meus agradecimentos.

O professor de Matemática, principalmente aquele que atua no chamado Segundo Grau, no escasso tempo que lhe resta da faina diária para preparar suas aulas, conta praticamente com uma única fonte de referência: o livro-texto que adota (ou os outros, que dele pouco diferem).

Visando dar ao professor maior apoio bibliográfico, a Sociedade Brasileira de Matemática, com a colaboração do IMPA, vem publicando na sua "Coleção do Professor de Matemática" uma série de monografias, cada uma delas dedicada a um tópico específico, principalmente a nível do Ensino Médio. A presente publicação, que pretende ser o primeiro livro de uma trilogia, tem a mesma finalidade. Só que agora, em vez de expor o programa de Matemática do segundo grau sob forma de temas isolados, estaremos dividindo os assuntos por série.

Em todo este livro, procuramos deixar claro que a Matemática oferece uma variedade de conceitos abstratos que servem de modelos para situações concretas, permitindo assim analisar, prever e tirar conclusões

de forma eficaz em circunstâncias onde uma abordagem empírica muitas vezes não conduz a nada. Todos os temas aqui abordados são apresentados dentro dessa ótica.

Assim é que os conjuntos são o modelo matemático para a organização do pensamento lógico; os números são o modelo para as operações de contagem e medida; as funções afins, as quadráticas, as exponenciais, as logarítmicas e as trigonométricas, cada uma delas é estudada como o modelo matemático adequado para representar uma situação específica.

A fim de saber qual o tipo de função que deve ser empregado para resolver um determinado problema, é necessário comparar as características desse problema com as propriedades típicas da função que se tem em mente. Este processo requer que se conheçam os teoremas de caracterização para cada tipo de função. Sem tal conhecimento é impossível aplicar satisfatoriamente os conceitos e métodos matemáticos para resolver os problemas concretos que ocorrem, tanto no dia-a-dia como nas aplicações da Matemática às outras ciências e à tecnologia.

Vários desses teoremas de caracterização são expostos aqui, de forma elementar. Acho que todos os professores devem conhecê-los e ensinar seus alunos a usá-los de forma consciente. Quanto às demonstrações desses teoremas, embora acessíveis, elas foram incluídas aqui para o entendimento dos professores. Não considero essencial repassá-las aos estudantes, salvo em casos especiais, a critério de cada professor.

O importante é ter em mente que as aplicações aqui sugeridas despertam o interesse, justificam o esforço, exibem a eficiência e a utilidade dos métodos da Matemática mas, por outro lado, só podem ser levadas a bom termo se contarem com uma base conceitual adequada.

A publicação deste livro contou com o apoio da FAPERJ, em convênio com a CAPES, com a valiosa e sempre presente colaboração do IMPA e com a proverbial expertise de Wilson Góes.

Rio de Janeiro, 26 de novembro, 1996

Elon Lages Lima

Prefácio à segunda edição

Esta edição difere da primeira apenas pela correção de alguns erros tipográficos, por uma pequena modificação no final do Capítulo 8, pela inclusão de dois novos exercícios no Capítulo 6 e pela eliminação de uma asneira que escrevi e que o Gugu me apontou.

Rio de Janeiro, julho, 1997
Elon Lages Lima

Prefácio à terceira edição

Nesta edição a seção 7 do Capítulo 6 foi re-escrita. Quero agradecer a Artur Ávila Cordeiro, por uma elegante sugestão ali incorporada. Agradeço também a Jonas Gomes pela imagem da página 146. Agradecimentos são devidos também a Maria Laura Magalhães Gomes por ter usado o texto num curso e ter feito uma cuidadosa revisão do mesmo.

Rio de Janeiro, dezembro, 1998
Elon Lages Lima

Prefácio à quinta edição

Para esta edição foram feitas algumas modificações de pequena monta, visando maior clareza e correção. Agradeço aos diversos colegas que me apontaram os defeitos, em particular ao Prof. Antonio Paiva, que fez uma revisão sistemática do texto.

Rio de Janeiro, outubro, 2000
Elon Lages Lima

Capítulo 1

Conjuntos

1. A Noção de Conjunto

Toda a Matemática atual é formulada na linguagem de conjuntos. Portanto, a noção de conjunto é a mais fundamental: a partir dela, todos os conceitos matemáticos podem ser expressos. Ela é também a mais simples das idéias matemáticas.

Um conjunto é formado por elementos. Dados um conjunto A e um objeto qualquer a (que pode até mesmo ser outro conjunto), a única pergunta cabível em relação a eles é: a é ou não um elemento do conjunto A? No caso afirmativo, diz-se que a *pertence* ao conjunto A e escreve-se $a \in A$. Caso contrário, põe-se $a \notin A$ e diz-se que a *não pertence* ao conjunto A.

A Matemática se ocupa primordialmente de números e do espaço. Portanto, os conjuntos mais freqüentemente encontrados na Matemática são os conjuntos numéricos, as figuras geométricas (que são conjuntos de pontos) e os conjuntos que se derivam destes, como os conjuntos de funções, de matrizes etc.

A linguagem dos conjuntos, hoje universalmente adotada na apresentação da Matemática, ganhou esta posição porque permite dar aos conceitos e às proposições desta ciência a precisão e a generalidade que constituem sua característica básica.

Os conjuntos substituem as "propriedades" e as "condições". Assim, em vez de dizermos que "o objeto x goza da propriedade P" ou "o objeto y satisfaz a condição C", podemos escrever $x \in A$ e $y \in B$, onde A é o conjunto dos objetos que gozam da propriedade

P e B é o conjunto dos objetos que satisfazem a condição C.

Por exemplo, sejam P a propriedade de um número inteiro x ser par (isto é, divisível por 2) e C a condição sobre o número real y expressa por

$$y^2 - 3y + 2 = 0.$$

Por outro lado, sejam

$$A = \{\ldots -4, -2, 0, 2, 4, 6, \ldots\} \quad \text{e} \quad B = \{1, 2\}.$$

Então, tanto faz dizer que x goza da propriedade P e y satisfaz a condição C como afirmar que $x \in A$ e $y \in B$.

Qual é, porém, a vantagem que se obtém quando se prefere dizer que $x \in A$ e $y \in B$ em vez de dizer que x goza da propriedade P e y satisfaz a condição C?

A vantagem de se utilizar a linguagem e a notação de conjuntos é que entre estes existe uma álgebra, montada sobre as operações de reunião ($A \cup B$) e interseção ($A \cap B$), além da relação de inclusão ($A \subset B$). As propriedades e regras operatórias dessa álgebra, como por exemplo

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{e} \quad A \subset A \cup B,$$

são extremamente fáceis de manipular e representam um enorme ganho em simplicidade e exatidão quando comparadas ao manuseio de propriedades e condições.

Recomendações:

1. Evite dizer "teoria dos conjuntos". Essa teoria existe mas, neste nível, está-se apenas introduzindo a linguagem e a notação dos conjuntos. Não há teoria alguma aqui.

2. Resista à tentação de usar a expressão "x satisfaz a propriedade P". Um objeto pode *gozar* de uma propriedade, *possuir* uma propriedade, ou *ter* uma propriedade. Pode também *satisfazer* uma condição ou *cumprir* essa condição. Satisfazer uma propriedade é tão errado como gozar de uma condição. Propriedade é sinônimo

de atributo; condição é o mesmo que requisito.

3. Nunca escreva coisas como $A = \{\text{conjunto dos números pares}\}$. Isto é incorreto. O símbolo $\{\ldots\}$ significa o conjunto cujos elementos estão descritos no interior das chaves. Escreva A = conjunto dos número pares, $A = \{\text{números pares}\}$ ou $A = \{2n; n \in \mathbb{Z}\}$.

Existe um conjunto excepcional e intrigante: o conjunto vazio, designado pelo símbolo $\varnothing$. Ele é aceito como conjunto porque cumpre a utilíssima função de simplificar as proposições, evitando uma longa e tediosa menção de exceções. Qualquer propriedade contraditória serve para definir o conjunto vazio. Por exemplo, tem-se $\varnothing = \{x; x \neq x\}$, ou seja, $\varnothing$ é o conjunto dos objetos x tais que x é diferente de si mesmo. Seja qual for o objeto x, tem-se sempre $x \notin \varnothing$. Em muitas questões matemáticas é importante saber que um determinado conjunto X não é vazio. Para mostrar que X não é vazio, deve-se simplesmente encontrar um objeto x tal que $x \in X$.

Outros conjuntos curiosos são os conjuntos unitários. Dado um objeto x qualquer, o conjunto unitário $\{x\}$ tem como único elemento esse objeto x. Estritamente falando, x e $\{x\}$ não são a mesma coisa. Por exemplo, $\varnothing \neq \{\varnothing\}$ pois $\{\varnothing\}$ possui um elemento (tem-se $\varnothing \in \{\varnothing\}$) mas $\varnothing$ é vazio. Em certas ocasiões, entretanto, pode tornar-se um pedantismo fazer essa distinção. Nesses casos, admite-se escrever x em vez de $\{x\}$. Um exemplo disso ocorre quando se diz que a interseção de duas retas r e s é o ponto P (em lugar do conjunto cujo único elemento é P) e escreve-se $r \cap s = P$, em vez de $r \cap s = \{P\}$. (Com experiência e bom senso, quem se ocupa de Matemática percebe que a obediência estrita aos rígidos padrões da notação e do rigor, quando praticada ao pé da letra, pode ser um obstáculo à clareza, à elegância e ao entendimento dos alunos.)

2. A Relação de Inclusão

Sejam A e B conjuntos. Se todo elemento de A for também elemento de B, diz-se que A é um *subconjunto* de B, que A está *con-*

tido em B ou que A é *parte* de B. Para indicar este fato, usa-se a notação $A \subset B$.

Exemplo: sejam T o conjunto dos triângulos e P o conjunto dos polígonos do plano. Todo triângulo é um polígono, logo $T \subset P$.

A relação $A \subset B$ chama-se *relação de inclusão*. Quando A não é um subconjunto de B, escreve-se $A \not\subset B$. Isto significa que nem todo elemento de A pertence a B, ou seja, que existe pelo menos um objeto a tal que $a \in A$ e $a \notin B$. Por exemplo, sejam A o conjunto dos números pares e B o conjunto dos múltiplos de 3. Tem-se $A \not\subset B$ porque $2 \in A$ mas $2 \notin B$. Tem-se também $B \not\subset A$ pois $3 \in B$ mas $3 \notin A$.

Há duas inclusões extremas. A primeira é óbvia: para todo conjunto A, vale $A \subset A$ (pois é claro que todo elemento de A pertence a A). A outra é, no mínimo, curiosa: tem-se $\varnothing \subset A$, seja qual for o conjunto A. Com efeito, se quiséssemos mostrar que $\varnothing \not\subset A$, teríamos que obter um objeto x tal que $x \in \varnothing$ mas $x \notin A$. Como $x \in \varnothing$ é impossível, somos levados a concluir que $\varnothing \subset A$, ou seja, que o conjunto vazio é subconjunto de qualquer outro.

Diz-se que A é um *subconjunto próprio* de B quando se tem $A \subset B$ com $A \neq \varnothing$ e $A \neq B$.

A relação de inclusão goza de três propriedades fundamentais. Dados quaisquer conjuntos A, B C tem-se:

reflexividade: $A \subset A$;

anti-simetria: se $A \subset B$ e $B \subset A$ então $A = B$;

transitividade: se $A \subset B$ e $B \subset C$ então $A \subset C$.

A propriedade anti-simétrica é constantemente usada nos raciocínios matemáticos. Quando se deseja mostrar que os conjuntos A e B são iguais, prova-se que $A \subset B$ e $B \subset A$, ou seja, que todo elemento de A pertence a B e todo elemento de B pertence a A. Na realidade, a propriedade anti-simétrica da relação de inclusão contém, nela embutida, a condição de igualdade entre conjuntos: os conjuntos A e B são iguais se, e somente se, têm os mesmos elementos.

Por sua vez, a propriedade transitiva da inclusão é a base do raciocínio dedutivo, sob a forma que classicamente se chama de *silogismo*. Um exemplo de silogismo (tipicamente aristotélico) é o seguinte: todo ser humano é um animal, todo animal é mortal, logo todo ser humano é mortal. Na linguagem de conjuntos, isso seria formulado assim: sejam H, A e M respectivamente os conjuntos dos seres humanos, dos animais e dos mortais. Temos $H \subset A$ e $A \subset M$, logo $H \subset M$.

Recomendações:

4. Se a é um elemento do conjunto A, a relação $a \in A$ pode também ser escrita sob a forma $\{a\} \subset A$. Mas é incorreto escrever $a \subset A$ e $\{a\} \in A$.

5. Em Geometria, uma reta, um plano e o espaço são conjuntos. Seus elementos são pontos. Se r é uma reta contida no plano Π, escreve-se $r \subset \Pi$ pois, neste caso, a reta r é um subconjunto do plano Π. Não se deve escrever $r \in \Pi$ nem dizer que a reta r pertence ao plano Π, pois os elementos do conjunto Π são pontos e não retas.

A relação de inclusão entre conjuntos está estreitamente relacionada com a implicação lógica. Vejamos como. Sejam P e Q propriedades referentes a um elemento genérico de um conjunto U. Essas propriedades definem os conjuntos A, formado pelos elementos de U que gozam de P, e B, conjunto formado pelos elementos de U que têm a propriedade Q. Diz-se então que a propriedade P *implica* (ou *acarreta*) a propriedade Q, e escreve-se $P \Rightarrow Q$, para significar que $A \subset B$.

Por exemplo, seja U o conjunto dos quadriláteros convexos do plano. Designemos com P a propriedade de um quadrilátero ter seus quatro ângulos retos e por Q a propriedade de um quadrilátero ter seus lados opostos paralelos. Então podemos escrever $P \Rightarrow Q$. Com efeito, neste caso, A é o conjunto dos retângulos e B é o conjunto dos paralelogramos, logo $A \subset B$.

Vejamos outro exemplo. Podemos escrever a implicação

$$x^2 + x - 1 = 0 \quad \Rightarrow \quad x^3 - 2x + 1 = 0.$$

Ela significa que toda raiz da equação $x^2 + x - 1 = 0$ é também raiz de

$$x^3 - 2x + 1 = 0.$$

Há diferentes maneiras de se ler a relação $P \Rightarrow Q$. Pode-se dizer "P implica Q", "se P então Q", "P é condição suficiente para Q", "Q é condição necessária para P" ou "P somente se Q".

Assim, no primeiro exemplo acima, podemos dizer: "ser retângulo implica ser paralelogramo", "se x é um retângulo então x é um paralelogramo", "ser retângulo é condição suficiente para ser paralelogramo", "ser paralelogramo é condição necessária para ser retângulo", ou, finalmente, "todo retângulo é um paralelogramo".

A implicação $Q \Rightarrow P$ chama-se a *recíproca* de $P \Rightarrow Q$. Evidentemente, a recíproca de uma implicação verdadeira pode ser falsa. Nos dois exemplos dados acima, as recíprocas são falsas: nem todo paralelogramo é retângulo e $x = 1$ é raiz da equação

$$x^3 - 2x + 1 = 0$$

mas não da equação

$$x^2 + x - 1 = 0.$$

Quando são verdadeiras ambas as implicações $P \Rightarrow Q$ e $Q \Rightarrow P$, escreve-se $P \Leftrightarrow Q$ e lê-se "P se, e somente se, Q", "P é equivalente a Q" ou "P é necessária e suficiente para Q". Isto significa que o conjunto dos elementos que gozam da propriedade P coincide com o conjunto dos elementos que gozam de Q.

Por exemplo, sejam P a propriedade de um triângulo, cujos lados medem $x \leqslant y \leqslant z$, ser retângulo e Q a propriedade de valer

$$z^2 = x^2 + y^2.$$

Então $P \Leftrightarrow Q$.

Recomendações:

6. Nunca escreva (ou diga) coisas do tipo

$$\text{“se} \quad x^2+x-1=0 \quad \Rightarrow \quad x^3-2x+1=0\text{”.}$$

O símbolo $\Rightarrow$ não significa "então", mas sim "implica". Também é incorreto empregar o símbolo $\Rightarrow$ com o significado conclusivo da palavra "portanto". O símbolo adequado para esta palavra é $\therefore$ e não $\Rightarrow$.

7. As definições matemáticas consistem em atribuir nomes a objetos que gozam de certas propriedades particularmente interessantes. Elas contribuem para a clareza do discurso e a economia do pensamento. Por exemplo, um número natural $n > 1$ chama-se primo quando 1 e n são os únicos números naturais que são seus divisores. Embora, estritamente falando, não seja errado usar "se, e somente se," numa definição, trata-se de um costume didaticamente inadequado pois dá a impressão de um teorema, além de ocultar o fato de que se trata de simplesmente dar um nome a um conceito. Por exemplo, se queremos definir *paralelogramo* devemos dizer assim: "chama-se paralelogramo a um quadrilátero no qual os lados opostos são paralelos". Alguns autores escrevem: "um quadrilátero é um paralelogramo se, e somente se, os lados opostos são paralelos". Isto não tem cara de definição.

Duas observações adicionais a respeito de proposições matemáticas:

A primeira é que em Matemática não há afirmações absolutas ou peremptórias. *Todas* as proposições matemáticas são do tipo "se P então Q". (Esta afirmação peremptória não pertence à Matemática. Ela é apenas sobre Matemática.)

Por exemplo, seja o Teorema de Pitágoras. Ele parece uma verdade absoluta mas na realidade é uma afirmação condicional:

"*Se* $a > b \geqslant c$ são as medidas dos lados de um triângulo retân-

gulo *então* $a^2 = b^2 + c^2$."

Por isso as vezes se diz que a Matemática é a ciência das condições necessárias. Ou então se diz como Bertrand Russel: "Na Matemática nunca sabemos do que estamos falando nem se é verdade o que estamos dizendo".

A segunda observação diz respeito às afirmações que são vacuamente satisfeitas. Se um professor disser à sua classe que todos os alunos que tiverem 5 metros de altura passarão com nota 10 sem precisar de prestar exames, ele certamente estará falando a verdade, mesmo que corrija suas provas com o máximo rigor. Com efeito, sejam P a propriedade de um aluno ter 5 metros de altura e Q a de obter nota 10 sem prestar exames. Então $P \Rightarrow Q$ pois o conjunto definido pela propriedade P é vazio e o conjunto vazio está contido em qualquer outro. De um modo geral, a implicação $P \Rightarrow Q$ é verdadeira (vacuamente) sempre que não haja elementos com a propriedade P.

Às vezes é mais natural dizer que um objeto cumpre uma certa *condição* em lugar de afirmar que ele possui uma determinada *propriedade*. Por exemplo, uma equação como $x^2 - x - 2 = 0$ é mais apropriadamente vista como uma condição a que deve satisfazer o número x do que uma propriedade desse número. (Estamos falando de "mais ou menos conveniente", não de "certo ou errado".)

A propósito, a resolução de uma equação é um caso típico em que se tem uma seqüência de implicações lógicas. Vejamos. Para resolver a equação

$$x^2 - x - 2 = 0$$

podemos seguir os passos abaixo:

(P) ...... $x^2 - x - 2 = 0$;

(Q) ...... $(x-2)(x+1) = 0$;

(R) ...... $x = 2$ ou $x = -1$;

(S) ...... $x \in \{2, -1\}$.

Se chamarmos respectivamente de P, Q, R e S as condições impostas sobre o número x em cada uma das linhas acima, os

passos que acabamos de seguir significam que

$$P \Rightarrow Q \Rightarrow R \Rightarrow S,$$

isto é, se o número x satisfaz P então satisfaz Q e assim por diante. Por transitividade, a conclusão a tirar é $P \Rightarrow S$, ou seja:

$$\text{Se} \quad x^2 - x - 2 = 0 \quad \text{então} \quad x \in \{2, -1\}.$$

Estritamente falando, esta afirmação não significa que as raízes da equação $x^2-x-2 = 0$ são 2 e -1. O que está dito acima é que se houver raízes desta equação elas devem pertencer ao conjunto $\{2, -1\}$. Acontece, entretanto, que no presente caso, os passos acima podem ser revertidos. É fácil ver que valem as implicações recíprocas $S \Rightarrow R \Rightarrow Q \Rightarrow P$, logo $S \Rightarrow P$. Portanto $P \Leftrightarrow S$, ou seja, 2 e -1 são de fato as (únicas) raízes da equação

$$x^2 - x - 2 = 0.$$

É importante, quando se resolve uma equação, ter em mente que cada passo do processo adotado representa uma implicação lógica. Às vezes essa implicação não pode ser revertida (isto é, sua recíproca não é verdadeira). Nesses casos, o conjunto obtido no final apenas contém (mas não é igual a) o conjunto das raízes, este último podendo até mesmo ser vazio. Ilustremos esta possibilidade com um exemplo.

Seja a equação $x^2+1 = 0$. Sabemos que ela não possui soluções reais. Na seqüência abaixo, cada uma das letras P, Q, R e S representa a condição sobre o número x expressa na igualdade ao lado. Assim, P significa $x^2 + 1 = 0$, etc.

(P) $x^2 + 1 = 0$. (multiplicando por $x^2 - 1$)

(Q) $x^4 - 1 = 0$;

(R) $x^4 = 1$;

(S) $x \in \{-1, 1\}$.

Evidentemente, tem-se $P \Rightarrow Q \Rightarrow R \Rightarrow S$, logo $P \Rightarrow S$, ou seja, toda raiz real da equação $x^2 + 1 = 0$ pertence ao conjunto $\{-1, 1\}$. O

raciocínio é absolutamente correto, mas apenas ilustra o fato de que o conjunto vazio está contido em qualquer outro. A conclusão que se pode tirar é que se houver raízes reais da equação $x^2 + 1 = 0$ elas pertencerão ao conjunto $\{-1, 1\}$. Nada mais. O fato é que a implicação $P \Rightarrow Q$ não pode ser revertida: sua recíproca é falsa. Este fenômeno ocorre freqüentemente quando se estudam as chamadas "equações irracionais", mas às vezes ele se manifesta de forma sutil, provocando perplexidade. (Veja Exercício 6.)

Observação:

Não é raro que pessoas confundam "necessário" com "suficiente". A.C.M. notou que os alunos têm mais facilidade de usar corretamente esta última palavra do que a anterior, já que "suficiente" é sinônimo de "bastante". Talvez isso tenha a ver com o fato de que uma condição suficiente é geralmente mais forte do que a conclusão a que se quer chegar. Por exemplo, para que um número seja par é suficiente que seja múltiplo de 4. (Ou basta ser múltiplo de 4 para ser par.) Por outro lado, uma condição necessária é, em geral, mais fraca do que a conclusão desejada. Assim, por exemplo, para que um quadrilátero convexo Q seja um retângulo é necessário que seus lados opostos sejam paralelos, mas esta propriedade apenas não assegura que Q tenha seus ângulos todos retos. É claro que um conjunto completo de condições necessárias para que seja válida uma propriedade P constitui uma condição suficiente para P.

3. O Complementar de um Conjunto

A noção de complementar de um conjunto só faz pleno sentido quando se fixa um conjunto U, chamado o *universo do discurso*, ou *conjunto-universo*. U poderia ser chamado o assunto da discussão, ou o tema em pauta: estaremos falando somente dos elementos de U.

Uma vez fixado U, todos os elementos a serem considerados pertencerão a U e todos os conjuntos serão subconjuntos de U, ou

derivados destes. Por exemplo: na Geometria Plana, U é o plano. Na teoria aritmética da divisibilidade, U é o conjunto dos números inteiros.

Então, dado um conjunto A (isto é, um subconjunto de U), chama-se *complementar* de A ao conjunto A^c formado pelos objetos de U que não pertencem a A. Lembremos que, fixado o conjunto A, para cada elemento x em U, vale uma, e somente uma, das alternativas: $x \in A$, ou $x \notin A$.

O fato de que, para todo $x \in U$, não existe uma outra opção além de $x \in A$ ou $x \notin A$ é conhecido em Lógica como o *princípio do terceiro excluído*, e o fato de que as alternativas $x \in A$ e $x \notin A$ não podem ser verdadeiras ao mesmo tempo chama-se o *princípio da não-contradição.*

Seguem-se dos princípios acima enunciados as seguintes regras operatórias referentes ao complementar:

(1) Para todo conjunto $A \subset U$, tem-se $(A^c)^c = A$. (Todo conjunto é o complementar do seu complementar.)

(2) Se $A \subset B$ então $B^c \subset A^c$. (Se um conjunto está contido noutro, seu complementar contém esse outro.)

A regra (2) pode ser escrita com a notação $\Rightarrow$, assumindo a forma seguinte

$$A \subset B \quad \Rightarrow \quad B^c \subset A^c.$$

Na realidade, na presença da regra (1), a regra (2) pode ser reforçada, valendo a equivalência abaixo

(3) $\quad A \subset B \Leftrightarrow B^c \subset A^c.$

Esta equivalência pode ser olhada sob o ponto de vista lógico, usando-se as propriedades P e Q que definem respectivamente os conjuntos A e B. Então o conjunto A é formado pelos elementos de U que gozam da propriedade P, enquanto que os elementos de B são todos os que (pertencem a U e) gozam da propriedade Q. As propriedades que definem os conjuntos A^c e B^c são respectivamente a *negação de* P, representada por P′, e a *negação de* Q, representada por Q′. Assim, dizer que um objeto x goza da propriedade P′ significa (por definição) afirmar que x não goza da propriedade P

(e analogamente, para Q'). Com estas convenções, a relação (3) acima se lê assim:

(4) $P \Rightarrow Q$ se, e somente se, $Q' \Rightarrow P'$.

Noutras palavras, a implicação $P \Rightarrow Q$ (P implica Q) equivale a dizer que $Q' \Rightarrow P'$ (a negação de Q implica a negação de P).

Vejamos um exemplo. Sejam U o conjunto dos quadriláteros convexos, R a propriedade que tem um quadrilátero x de ser um retângulo e P a propriedade de ser um paralelogramo. Então P' é a propriedade que tem um quadrilátero convexo de não ser um paralelogramo e R' a de não ser um retângulo. As implicações $R \Rightarrow P$ e $P' \Rightarrow R'$ se lêem, neste caso, assim:

(a) Se x é um retângulo então x é um paralelogramo;

(b) Se x não é um paralelogramo então x não é um retângulo.

Evidentemente, as afirmações (a) e (b) são equivalentes, ou seja, elas são apenas duas maneiras diferentes de dizer a mesma coisa.

A implicação $Q' \Rightarrow P'$ chama-se a *contrapositiva* da implicação $P \Rightarrow Q$.

Sob o ponto de vista pragmático, a contrapositiva de uma implicação nada mais é do que a mesma implicação dita com outras palavras, ou vista de um ângulo diferente. Assim, por exemplo, a afirmação de que todo número primo maior do que 2 é ímpar e a afirmação de que um número par maior do que 2 não é primo dizem exatamente a mesma coisa, ou seja, exprimem a mesma idéia, só que com diferentes termos.

No dia-a-dia da Matemática é freqüente, e muitas vezes útil, substituir uma implicação por sua contrapositiva, a fim de tornar seu significado mais claro ou mais manejável. Por isso é extremamente importante entender que $P \Rightarrow Q$ e $Q' \Rightarrow P'$ são afirmações equivalentes.

A equivalência entre uma implicação e sua contrapositiva é a base das *demonstrações por absurdo*.

Vejamos um exemplo.

No plano Π, consideremos as retas perpendiculares r e s. Seja

P a propriedade que tem uma reta x, nesse mesmo plano, de ser diferente de s e perpendicular a r. Por outro lado, seja Q a propriedade de uma reta x (ainda no plano Π) ser paralela a s. Então P', negação de P, é a propriedade de uma reta em Π coincidir com s ou não ser perpendicular a r. A negação de Q é a propriedade Q' que tem uma reta do plano Π de não ser paralela a s.

A implicação $P \Rightarrow Q$ se lê, em linguagem comum, assim: se duas retas distintas (s e x) são perpendiculares a uma terceira (a saber, r) então elas (s e x) são paralelas.

A contrapositiva $Q' \Rightarrow P'$ significa: se duas retas distintas não são paralelas então elas não são perpendiculares a uma terceira.

(Nos dois parágrafos acima estamos tratando de retas do mesmo plano.)

Acontece que é mais fácil (e mais natural) provar a implicação $Q' \Rightarrow P'$ do que $P \Rightarrow Q$. Noutras palavras, prova-se que $P \Rightarrow Q$ por absurdo. O raciocínio é bem simples: se as retas distintas s e x não são paralelas elas têm um ponto A em comum. Então, como é única a perpendicular s à reta r pelo ponto A, segue-se que x não é perpendicular a r.

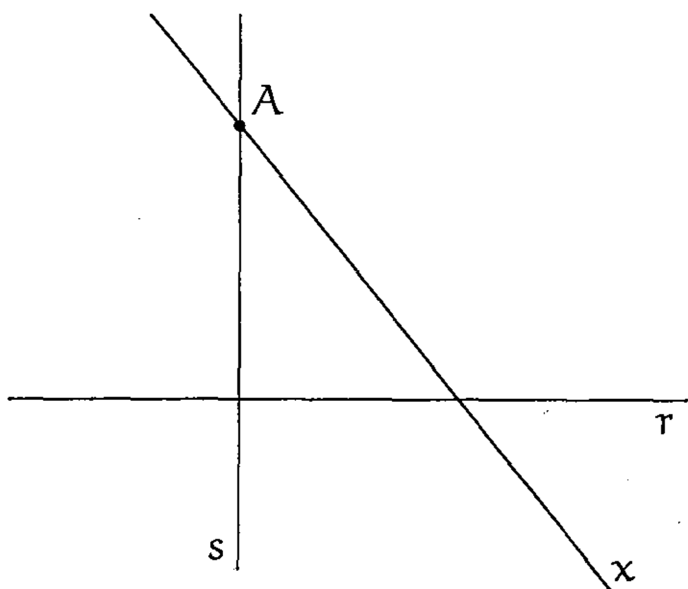


Figura 1

Observação:

Para provar que duas retas são paralelas, em geral se usa a demonstração por absurdo pois a definição de retas paralelas é baseada numa negação. (Retas paralelas são retas coplanares que *não* possuem pontos em comum.)

Observemos que se U é o universo então $U^c = \varnothing$ e $\varnothing^c = U$.

Recomendação:

8. Muitas vezes (principalmente nos raciocínios por absurdo) é necessário negar uma implicação $P \Rightarrow Q$. É preciso ter cuidado ao fazer isto. A negação de "todo homem é mortal" não é "nenhum homem é mortal" mas "existe (pelo menos) um homem imortal". Mais geralmente, negar que $P \Rightarrow Q$ significa admitir que existe (pelo menos) um objeto que tem a propriedade P mas não tem a propriedade Q. Isto é bem diferente de admitir que nenhum objeto com a propriedade P tem também a propriedade Q. Por exemplo, se P é a propriedade que tem um triângulo de ser isósceles e Q a propriedade de ser equilátero, a implicação $P \Rightarrow Q$ significaria que todo triângulo isósceles é equilátero (o que é falso). A negação de $P \Rightarrow Q$ é a afirmação de que existe (pelo menos) um triângulo isósceles não-equilátero.

Neste contexto, convém fazer uma distinção cuidadosa entre a idéia matemática de *negação* e a noção (não-matemática) de *contrário*, ou *oposto*. Se um conceito é expresso por uma palavra, o conceito contrário é expresso pelo antônimo daquela palavra. Por exemplo, o contrário de gigantesco é minúsculo, mas a negação de gigantesco inclui outras gradações de tamanho além de minúsculo.

4. Reunião e Interseção

Dados os conjuntos A e B, a *reunião* $A \cup B$ é o conjunto formado pelos elementos de A mais os elementos de B, enquanto que a *interseção* $A \cap B$ é o conjunto dos objetos que são ao mesmo tempo

elementos de A e de B. Portanto, se consideramos as afirmações

$$x \in A, \qquad x \in B,$$

veremos que $x \in A \cup B$ quando *pelo menos uma* dessas afirmações for verdadeira e, por outro lado, $x \in A \cap B$ quando *ambas* as afirmações acima forem verdadeiras.

Mais concisamente:

$$x \in A \cup B \quad \text{significa} \quad \text{“}x \in A \text{ ou } x \in B\text{”};$$

$$x \in A \cap B \quad \text{significa} \quad \text{“}x \in A \text{ e } x \in B\text{”}.$$

Nota-se, deste modo, que as operações $A \cup B$ e $A \cap B$ entre conjuntos constituem a contrapartida matemática dos conectivos lógicos “ou” e “e”. Assim, quando o conjunto A é formado pelos elementos que gozam da propriedade P e B pelos que gozam da propriedade Q então a propriedade que define o conjunto $A \cup B$ é “P ou Q” e o conjunto $A \cap B$ é definido pela propriedade “P e Q”.

Por exemplo, convencionemos dizer que um número x goza da propriedade P quando valer a igualdade

$$x^2 - 3x + 2 = 0.$$

Digamos ainda que x tem a propriedade Q quando for

$$x^2 - 5x + 6 = 0.$$

O conjunto dos números que possuem a propriedade P é $A = \{1, 2\}$ e o conjunto dos números que gozam de Q é $B = \{2, 3\}$. Assim, a afirmação

$$\text{“}x^2 - 3x + 2 = 0 \quad \text{ou} \quad x^2 - 5x + 6 = 0\text{”}$$

equivale a

$$\text{“}x \in \{1, 2, 3\},\text{”}$$

e a afirmação

$$\text{“}x^2 - 3x + 2 = 0 \quad \text{e} \quad x^2 - 5x + 6 = 0\text{”}$$

equivale a

$$\text{“}x \in \{2\} \quad \text{ou} \quad x = 2.\text{”}$$

Noutras palavras,

$$A \cup B = \{1, 2, 3\} \quad \text{e} \quad A \cap B = \{2\}.$$

É importante ressaltar que a palavra “ou” em Matemática tem um significado específico um tanto diferente daquele que lhe é atribuído na linguagem comum. No dia-a-dia, “ou” quase sempre liga duas alternativas incompatíveis (“vamos de ônibus ou de trem?”). Em Matemática, a afirmação “P ou Q” significa que pelo menos uma das alternativas P ou Q é válida, podendo perfeitamente ocorrer que ambas sejam. Por exemplo, é correta a afirmação “todo número inteiro é maior do que 10 ou menor do que 20”. Noutras palavras, se

$$A = \{x \in \mathbb{Z}; x > 10\}$$

e

$$B = \{x \in \mathbb{Z}; x < 20\}$$

então $A \cup B = \mathbb{Z}$.

A diferença entre o uso comum e o uso matemático do conectivo “ou” é ilustrada pela anedota do obstetra que também era matemático. Ao sair da sala onde acabara de realizar um parto, foi abordado pelo pai da criança, que lhe perguntou: “Foi menino ou menina, doutor?”. Resposta do médico: “Sim.” (Com efeito, se A é o conjunto das meninas, B o conjunto dos meninos e x o recém-nascido, certamente tem-se $x \in A \cup B$.)

As operações de reunião e intersecção são obviamente comutativas

$$A \cup B = B \cup A \quad \text{e} \quad A \cap B = B \cap A,$$

e associativas

$$(A \cup B) \cup C = A \cup (B \cup C)$$

e

$$(A \cap B) \cap C = A \cap (B \cap C).$$

Além disso, cada uma delas é distributiva em relação à outra:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

e

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Estas igualdades, que podem ser verificadas mediante a consideração dos casos possíveis, constituem, na realidade, regras que regem o uso combinado dos conectivos lógicos "ou" e "e".

A conexão entre as operações $\cup$, $\cap$ e a relação de inclusão $\subset$ é dada pelas seguintes equivalências:

$$A \cup B = B \quad \Leftrightarrow \quad A \subset B \quad \Leftrightarrow \quad A \cap B = A.$$

Além disso $A \subset B \Rightarrow A \cup C \subset B \cup C$ e $A \cap C \subset B \cap C$ para todo C.

E, finalmente, se A e B são subconjuntos do universo $\mathcal{U}$, tem-se:

$$(A \cup B)^c = A^c \cap B^c \quad \text{e} \quad (A \cap B)^c = A^c \cup B^c.$$

Estas relações, atribuídas ao matemático inglês Augustus de Morgan, significam que a negação de "P ou Q" é "nem P nem Q" e a negação de "P e Q" é "não P ou não Q".

5. Comentário Sobre a Noção de Igualdade

Uma coisa só é igual a si própria.

Quando se escreve $a = b$, isto significa que a e b são símbolos usados para designar o mesmo objeto.

Por exemplo, se a é a reta perpendicular ao segmento AB, levantada a partir do seu ponto médio e b é o conjunto dos pontos do plano que são equidistantes de A e B então $a = b$.

Em Geometria, às vezes ainda se usam expressões como "os ângulos α e β são iguais" ou "os triângulos ABC e $A'B'C'$ são iguais" para significar que são figuras que podem ser superpostas exatamente uma sobre a outra. A rigor, porém, esta terminologia é inadequada. Duas figuras geométricas que coincidem por superposição devem ser chamadas *congruentes*.

Talvez valha a pena observar que a palavra "igual" em Geometria já foi usada num sentido até bem mais amplo. Euclides, que viveu há 2300 anos, chamava "iguais" a dois segmentos de reta com o mesmo comprimento, a dois polígonos com a mesma área e a dois sólidos com o mesmo volume.

Na linguagem corrente, às vezes se diz que duas pessoas ou objetos são iguais quando um certo atributo, ao qual se refere o discurso naquele momento, é possuído igualmente pelas pessoas ou objetos em questão. Assim, por exemplo, quando dizemos que "todos são iguais perante a lei", isto significa que dois cidadãos quaisquer têm os mesmos direitos e deveres legais.

A relação "a é igual a b", que se escreve $a = b$, goza das seguintes propriedades:

Reflexividade: $a = a$;

Simetria: se $a = b$ então $b = a$;

Transitividade: se $a = b$ e $b = c$ então $a = c$.

Diante da simetria, a transitividade também se exprime assim: se $a = b$ e $c = b$ então $a = c$. Em palavras: dois objetos (a e c) iguais a um terceiro (b) são iguais entre si. Formulada deste modo, esta propriedade era uma das *noções comuns* (ou axiomas) que Euclides enunciou nas primeiras páginas do seu famoso livro "Os Elementos".

6. Recomendações Gerais

A adoção da linguagem e da notação de conjuntos em Matemática só se tornou uma prática universal a partir da terceira ou quarta década do século vinte. Esse uso, responsável pelos elevados graus de precisão, generalidade e clareza nos enunciados, raciocínios e definições, provocou uma grande revolução nos métodos, no alcance e na profundidade dos resultados matemáticos. No final do século 19, muitos matemáticos ilustres viam com séria desconfiança as novas idéias lançadas nos trabalhos pioneiros de G. Cantor. Mas, lenta e seguramente, esse ponto de vista se impôs e, no dizer

de D. Hilbert, com sua extraordinária autoridade, "ninguém nos expulsará desse paraíso que Cantor nos doou".

Portanto, se queremos iniciar os jovens em Matemática, é necessário que os familiarizemos com os rudimentos da linguagem e da notação dos conjuntos. Isto, inclusive, vai facilitar nosso próprio trabalho, pois a precisão dos conceitos é uma ajuda indispensável para a clareza das idéias. Mas, na sala de aula, há alguns cuidados a tomar. O principal deles refere-se ao comedimento, ao equilíbrio, à moderação. Isto consiste em evitar o pedantismo e os exageros que conduziram ao descrédito da onda de "Matemática Moderna". Não convém insistir em questões do tipo $\{\varnothing\} \neq \{\{\varnothing\}\}$ ou mesmo naquele exemplo $\varnothing \neq \{\varnothing\}$ dado acima.

Procure, sempre que possível, ilustrar seus conceitos com exemplos de conjuntos dentro da Matemática. Além de contribuir para implantar a linguagem de conjuntos, este procedimento pode também ajudar a relembrar, ou até mesmo aprender, fatos interessantes sobre Geometria, Aritmética, etc.

Seja cuidadoso, a fim de evitar cometer erros. A auto-crítica é o maior aliado do bom professor. Em cada aula, trate a si mesmo como um aluno cujo trabalho está sendo examinado. Pense antes no que vai dizer mas critique-se também depois: será que falei bobagem? Se achar que falou, não hesite em corrigir-se em público. Longe de desprestigiar, esse hábito fortalecerá a confiança dos alunos no seu mestre.

Esteja atento também à correção gramatical. Linguagem correta é essencial para a limpidez do raciocínio. Muitos dos nossos colegas professores de Matemática, até mesmo autores de livros, são um tanto descuidados a esse respeito. Dizem, por exemplo que "a reta r *intercepta* o plano α no ponto P", quando deveriam dizer *intersecta* (ou interseta) já que o ponto P é a interseção (ou intersecção) mas não a interceptação de r com α.

Eis aqui outros erros comuns de linguagem que devem ser evitados:

"Maior ou igual a". O correto é: "maior do que ou igual a".

(Tente dizer "igual ou maior a" e veja como soa mal.)

"Euclideano". O correto é "euclidiano".

"Assumir", no lugar de "supor" (vamos assumir que as retas r e s sejam paralelas). Isto é correto em inglês mas não em português.

Não diga "completude", diga "completeza". (Belo $\rightarrow$ beleza; rico $\rightarrow$ riqueza; nobre $\rightarrow$ nobreza; completo $\rightarrow$ completeza.)

Não diga "Espaço de tempo". Espaço e tempo são conceitos físicos fundamentais e independentes. Não se deve misturá-los. Diga "intervalo de tempo".

Exercícios

1. Sejam P_1, P_2, Q_1, Q_2 propriedades referentes a elementos de um conjunto-universo U. Suponha que P_1 e P_2 esgotam todos os casos possíveis (ou seja, um elemento qualquer de U ou tem a propriedade P_1 ou tem P_2). Suponha ainda que Q_1 e Q_2 são incompatíveis (isto é, excluem-se mutuamente). Suponha, finalmente, que $P_1 \Rightarrow Q_1$ e $P_2 \Rightarrow Q_2$. Prove que valem as recíprocas: $Q_1 \Rightarrow P_1$ e $Q_2 \Rightarrow P_2$.

2. Enquadre no contexto do exercício anterior o seguinte fato geométrico: *Duas oblíquas que se afastam igualmente do pé da perpendicular são iguais. Se se afastam desigualmente então são desiguais e a maior é a que mais se afasta.*

3. Sejam $X_1 X_2$, $Y_1 Y_2$ subconjuntos do conjunto-universo U. Suponha que $X_1 \cup X_2 = U$ e $Y_1 \cap Y_2 = \varnothing$, que $X_1 \subset Y_1$ e que $X_2 \subset Y_2$. Prove que $X_1 = Y_1$ e $X_2 = Y_2$.

4. Compare o exercício anterior com o primeiro em termos de clareza e simplicidade dos enunciados. Mostre que qualquer um deles pode ser resolvido usando o outro. Estabeleça resultados análogos com n propriedades ou n subconjuntos em vez de 2. Veja no livro "Coordenadas no Espaço", (Coleção do Professor de Matemática, S.B.M.) pág. 83 uma utilização deste fato com $n = 8$.

5. Ainda no tema do primeiro exercício, seria válido substituir

as implicações $P_1 \Rightarrow Q_1$ e $P_2 \Rightarrow Q_2$ na hipótese por suas recíprocas $Q_1 \Rightarrow P_1$ e $Q_2 \Rightarrow P_2$?

6. Escreva as implicações lógicas que correspondem à resolução da equação $\sqrt{x} + 2 = x$, veja quais são reversíveis e explique o aparecimento de raízes estranhas. Faça o mesmo com a equação $\sqrt{x} + 3 = x$.

7. Mostre que, para todo $m > 0$, a equação $\sqrt{x} + m = x$ tem exatamente uma raiz.

8. Considere as seguintes (aparentes) equivalências lógicas:

$$\begin{aligned} x = 1 \quad &\Leftrightarrow \quad x^2 - 2x + 1 = 0 \\ &\Leftrightarrow \quad x^2 - 2 \cdot 1 + 1 = 0 \\ &\Leftrightarrow \quad x^2 - 1 = 0 \\ &\Leftrightarrow \quad x = \pm 1. \end{aligned}$$

Conclusão (?): $x = 1 \quad \Leftrightarrow \quad x = \pm 1$. Onde está o erro?

9. As raízes do polinômio $x^3 - 6x^2 + 11x - 6$ são 1, 2 e 3. Substitua, nesse polinômio, o termo $11x$ por $11 \times 2 = 22$, obtendo então $x^3 - 6x^2 + 16$, que ainda tem 2 como raiz mas não se anula para $x = 1$ nem $x = 3$. Enuncie um resultado geral que explique este fato e o relacione com o exercício anterior.

10. Expressões tais como "para todo" e "qualquer que seja" são chamadas de quantificadores e aparecem em sentenças dos tipos:

(1) "Para todo x, é satisfeita a condição $P(x)$"

(2) "Existe algum x que satisfaz a condição $P(x)$",

onde $P(x)$ é uma condição envolvendo a variável x.

a) Sendo A o conjunto de todos os objetos x (de um certo conjunto-universo U) que satisfazem a condição $P(x)$, escreva as sentenças (1) e (2) acima, usando a linguagem de conjuntos.

b) Quais são as negações de (1) e (2)? Escreva cada uma destas negações usando conjuntos e compare com as sentenças obtidas em a).

c) Para cada sentença abaixo, diga se ela é verdadeira ou falsa e forme sua negação:

- Existe um número real x tal que $x^2 = -1$.
- Para todo número inteiro n, vale $n^2 > n$.
- Para todo número real x, tem-se $x > 1$ ou $x^2 < 1$.
- Para todo número real x existe um número natural n tal que $n > x$.
- Existe um número natural n tal que, para todo número real x, tem-se $n > x$.

11. Considere os conjuntos abaixo:
F = conjunto de todos os filósofos
M = conjunto de todos os matemáticos
C = conjunto de todos os cientistas
P = conjunto de todos os professores

a) Exprima cada uma das afirmativas abaixo usando a linguagem de conjuntos:

1) Todos os matemáticos são cientistas.

2) Alguns matemáticos são professores.

3) Alguns cientistas são filósofos.

4) Todos os filósofos são cientistas ou professores.

5) Nem todo professor é cientista.

b) Faça o mesmo com as afirmativas abaixo:

6) Alguns matemáticos são filósofos.

7) Nem todo filósofo é cientista.

8) Alguns filósofos são professores.

9) Se um filósofo não é matemático, ele é professor.

10) Alguns filósofos são matemáticos.

c) Tomando as cinco primeiras afirmativas como hipóteses, verifique quais das afirmativas do segundo grupo são necessariamente verdadeiras.

12. O artigo 34 da Constituição Brasileira de 1988 diz o seguinte:

"A União não intervirá nos Estados nem no Distrito Federal, exceto para:

I. Manter a integridade nacional;

II. Repelir invasão estrangeira ou de unidade da Federação em outra"

III. .... ;

a) Suponhamos que o estado do Rio de Janeiro seja invadido por tropas do estado de São Paulo. O texto acima obriga a União a intervir no estado? Na sua opinião, qual era a intenção dos legisladores nesse caso?

b) Reescreva o texto do artigo 34 de modo a torná-lo mais preciso.

13. Prove que $x^2 + x - 1 = 0 \Rightarrow x^3 - 2x + 1 = 0$.

14. Prove que, para x, y, k inteiros, tem-se $x + 4y = 13k \iff 4x + 3y = 13(4k - y)$. Conclua que $4x + 3y$ e $x + 4y$ são divisíveis por 13 para os mesmos valores inteiros de x e y.

15. O diagrama de Venn para os conjuntos X, Y, Z decompõe o plano em oito regiões. Numere essas regiões e exprima cada um dos conjuntos abaixo como reunião de algumas dessas regiões. (Por exemplo: $X \cap Y = 1 \cup 2$.)

a) $(X^c \cup Y)^c$; b) $(X^c \cup Y) \cup Z^c$;
c) $(X^c \cap Y) \cup (X \cap Z^c)$; d) $(X \cup Y)^c \cap Z$

16. Exprimindo cada membro como reunião de regiões numeradas, prove as igualdades:

a) $(X \cup Y) \cap Z = (X \cap Z) \cup (Y \cap Z)$;
b) $X \cup (Y \cap Z)^c = X \cup Y^c \cup Z^c$.

17. Sejam A, B e C conjuntos. Determine uma condição necessária e suficiente para que se tenha $A \cup (B \cap C) = (A \cup B) \cap C$.

18. A diferença entre conjuntos é definida por $A - B = \{x \mid x \in A \text{ e } x \notin B\}$. Determine uma condição necessária e suficiente para que se tenha $A - (B - C) = (A - B) - C$.

19. Prove que se um quadrado perfeito é par então sua raiz quadrada é par e que se um quadrado perfeito é ímpar então sua raiz quadrada é ímpar.

20. Prove o teorema de Cantor: se A é um conjunto e $P(A)$ é o conjunto das partes de A, não existe uma função $f\colon A \to P(A)$ que seja sobrejetiva.
Sugestão: Suponha que exista uma tal função f e considere $X = \{x \in A : x \notin f(x)\}$.

Capítulo 2

Números Naturais

"Deus criou os números naturais. O resto é obra dos homens."
Leopold Kronecker

1. Introdução

Enquanto os conjuntos constituem um meio auxiliar, os números são um dos dois objetos principais de que se ocupa a Matemática. (O outro é o espaço, junto com as figuras geométricas nele contidas.)

Números são entes abstratos, desenvolvidos pelo homem como modelos que permitem contar e medir, portanto avaliar as diferentes quantidades de uma grandeza.

Os compêndios tradicionais dizem o seguinte:

"Número é o resultado da comparação entre uma grandeza e a unidade. Se a grandeza é discreta, essa comparação chama-se uma *contagem* e o resultado é um número inteiro; se a grandeza é contínua, a comparação chama-se uma *medição* e o resultado é um número real."

Nos padrões atuais de rigor matemático, o trecho acima não pode ser considerado como uma definição matemática, pois faz uso de idéias (como grandeza, unidade, discreta, contínua) e processos (como comparação) de significado não estabelecido. Entretanto, todas as palavras que nela aparecem possuem um sentido bastante claro na linguagem do dia-a-dia. Por isso, embora não sirva para demonstrar teoremas a partir dela, a definição tradicional tem o grande mérito de nos revelar para que servem e por qual motivo foram inventados os números. Isto é muito mais do que se pode

dizer sobre a definição que encontramos no nosso dicionário mais conhecido e festejado, conforme reproduzimos a seguir.

Número. [Do lat. *numeru.*] *S.m.* 1. *Mat.* O conjunto de todos os conjuntos equivalentes a um conjunto dado.

Discutiremos este ponto logo mais, quando tratarmos de números cardinais. No momento, parece oportuno fazer uma pequena pausa para uma observação.

2. Comentário: Definições, Axiomas, etc.

Conforme dissemos no Capítulo 1, uma definição matemática é uma convenção que consiste usar um nome, ou uma sentença breve, para designar um objeto ou uma propriedade, cuja descrição normalmente exigiria o emprego de uma sentença mais longa. Vejamos algumas definições, como exemplo:

- *Ângulo* é a figura formada por duas semi-retas que têm a mesma origem.
- *Primos entre si* são dois ou mais números naturais cujo único divisor comum é a unidade.

Mas nem sempre foi assim. Euclides, por exemplo, começa os "Elementos" com uma série de definições, das quais selecionamos as seguintes:

- *Linha* é um comprimento sem largura.
- *Superfície* é o que possui comprimento e largura somente.
- Quando uma reta corta outra formando ângulos adjacentes iguais, cada um desses ângulos chama-se *reto* e as retas se dizem *perpendiculares*.

As definições de ângulo e de números primos entre si, dadas acima, bem como as definições de ângulo reto e retas perpendiculares dadas por Euclides, são corretas. Elas atendem aos padrões atuais de precisão e objetividade. Por outro lado, nas definições de linha e superfície, Euclides visa apenas oferecer ao seu leitor uma imagem intuitiva desses conceitos. Elas podem servir para

ilustrar o pensamento geométrico mas não são utilizáveis nos raciocínios matemáticos porque são formuladas em termos vagos e imprecisos.

Na apresentação de uma teoria matemática, toda definição faz uso de termos específicos, os quais foram definidos usando outros termos, e assim sucessivamente. Este processo iterativo leva a três possibilidades:

a) Continua indefinidamente, cada definição dependendo de outras anteriores, sem nunca chegar ao fim.

b) Conduz a uma circularidade, como nos dicionários. (Onde se vê, por exemplo: compreender $\rightarrow$ perceber, perceber $\rightarrow$ entender e entender $\rightarrow$ compreender.)

c) Termina numa palavra, ou num conjunto de palavras (de preferência dotadas de conotações intuitivas simples) que não são definidas, isto é, que são tomadas como representativas de conceitos primitivos. Exemplos: ponto, reta, conjunto.

Evidentemente, as alternativas a) e b) acima citadas não convêm à Matemática. A alternativa c) é a adotada. Se prestarmos atenção, veremos que foi assim que aprendemos a falar. Numerosas palavras nos foram apresentadas sem definição e permanecem até hoje em nosso vocabulário como conceitos primitivos, que aprendemos a usar por imitação e experiência.

Para poder empregar os conceitos primitivos adequadamente, é necessário dispor de um conjunto de princípios ou regras que disciplinem sua utilização e estabeleçam suas propriedades. Tais princípios são chamados *axiomas* ou *postulados*. Assim como os conceitos primitivos são objetos que não se definem, os axiomas são proposições que não se demonstram.

Uma vez feita a lista dos conceitos primitivos e enunciados os axiomas de uma teoria matemática, todas as demais noções devem ser definidas e as afirmações seguintes devem ser demonstradas. Nisto consiste o chamado *método axiomático*. As proposições a serem demonstradas chamam-se *teoremas* e suas conseqüências

imediatas são denominadas *corolários*. Uma proposição auxiliar, usada na demonstração de um teorema, é chamada um *lema*.

Ser um axioma ou ser um teorema não é uma característica intrínseca de uma proposição. Dependendo da preferência de quem organiza a apresentação da teoria, uma determinada proposição pode ser adotada como axioma ou então provada como teorema, a partir de outra proposição que a substituiu na lista dos axiomas.

Na seção seguinte, veremos um resumo da teoria matemática dos números naturais, onde os conceitos primitivos são "número natural" e "sucessor" e os axiomas são os de Peano.

Do ponto de vista do ensino a nível do segundo grau, não tem cabimento expor a Matemática sob forma axiomática. Mas é necessário que o professor saiba que ela pode ser organizada sob a forma acima delineada. Uma linha de equilíbrio a ser seguida na sala de aula deve basear-se nos seguintes preceitos:

1. Nunca dar explicações falsas sob o pretexto de que os alunos ainda não têm maturidade para entender a verdade. (Isto seria como dizer a uma criança que os bebês são trazidos pela cegonha.) Exemplo: "infinito é um número muito grande". Para outro exemplo, vide RPM 29, págs. 13–19.

2. Não insistir em detalhes formais para justificar afirmações que, além de verdadeiras, são intuitivamente óbvias e aceitas por todos sem discussão nem dúvidas. Exemplo: o segmento de reta que une um ponto interior a um ponto exterior de uma circunferência tem exatamente um ponto em comum com essa circunferência.

Em contraposição, fatos importantes cuja veracidade não é evidente, como o Teorema de Pitágoras ou a Fórmula de Euler para poliedros convexos, devem ser demonstrados (até mesmo de várias formas diferentes).

Excetuam-se, naturalmente, demonstrações longas, elaboradas ou que façam uso de noções e resultados acima do alcance dos estudantes desse nível (como o Teorema Fundamental da Álgebra,

por exemplo).

Provar o óbvio transmite a falsa impressão de que a Matemática é inútil. Por outro lado, usar argumentos elegantes e convincentes para demonstrar resultados inesperados é uma maneira de exibir sua força e sua beleza. As demonstrações, quando objetivas e bem apresentadas, contribuem para desenvolver o raciocínio, o espírito crítico, a maturidade e ajudam a entender o encadeamento lógico das proposições matemáticas.

3. Ter sempre em mente que, embora a Matemática possa ser cultivada por si mesma, como um todo coerente, de elevado padrão intelectual, formado por conceitos e proposições de natureza abstrata, sua presença no currículo escolar não se deve apenas ao valor dos seus métodos para a formação mental dos jóvens.

A importância social da Matemática provém de que ela fornece *modelos* para analisar situações da vida real. Assim, por exemplos, conjuntos são o modelo para disciplinar o raciocínio lógico, números naturais são o modelo para contagem e números reais são o modelo para medida; funções afins servem de modelo para situações, como o movimento uniforme, em que os acréscimos da função são proporcionais aos acréscimos da variável independente. E assim por diante.

Todos os tópicos deste livro são abordados sob o seguinte lema: a Matemática fornece modelos abstratos para serem utilizados em situações concretas, do dia-a-dia e das Ciências. Para poder empregar estes modelos é necessário verificar, em cada caso, que as hipóteses que lhe servem de base são satisfeitas.

3. O Conjunto dos Números Naturais

Lentamente, à medida em que se civilizava, a humanidade apoderou-se desse modelo abstrato de contagem (um, dois, três, quatro, ...) que são os números naturais. Foi uma evolução demorada. As tribos mais rudimentares contam apenas *um, dois, muitos*. A língua inglesa ainda guarda um resquício desse estágio na palavra *thrice*, que tanto pode significar "três vezes" como "muito" ou "ex-

tremamente".

As necessidades provocadas por um sistema social cada vez mais complexo e as longas reflexões, possíveis graças à disponibilidade de tempo trazida pelo progresso econômico, conduziram, através dos séculos, ao aperfeiçoamento do extraordinário instrumento de avaliação que é o conjunto dos números naturais.

Decorridos muitos milênios, podemos hoje descrever concisa e precisamente o conjunto $\mathbb{N}$ dos números naturais, valendo-nos da notável síntese feita pelo matemático italiano Giuseppe Peano no limiar do século 20.

$\mathbb{N}$ é um conjunto, cujos elementos são chamados *números naturais*. A essência da caracterização de $\mathbb{N}$ reside na palavra "sucessor". Intuitivamente, quando $n, n' \in \mathbb{N}$, dizer que n' é o *sucessor* de n significa que n' vem logo depois de n, não havendo outros números naturais entre n e n'. Evidentemente, esta explicação apenas substitui "sucessor" por "logo depois", portanto não é uma definição. O termo primitivo "sucessor" não é definido explicitamente. Seu uso e suas propriedades são regidos por algumas regras, abaixo enumeradas:

a) Todo número natural tem um único sucessor;

b) Números naturais diferentes têm sucessores diferentes;

c) Existe um único número natural, chamado *um* e representado pelo símbolo 1, que não é sucessor de nenhum outro;

d) Seja X um conjunto de números naturais (isto é, $X \subset \mathbb{N}$). Se $1 \in X$ e se, além disso, o sucessor de todo elemento de X ainda pertence a X, então $X = \mathbb{N}$.

As afirmações a), b), c) e d) acima são conhecidas como os *axiomas de Peano*. Tudo o que se sabe sobre os números naturais pode ser demonstrado como conseqüência desses axiomas.

Um engenhoso processo, chamado *sistema de numeração decimal*, permite representar todos os números naturais com o auxílio dos símbolos 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9. Além disso, os primeiros números naturais têm nomes: o sucessor do número um chama-

se "dois", o sucessor de dois chama-se "três", etc. A partir de um certo ponto, esses nomes tornam-se muito complicados, sendo preferível abrir mão deles e designar os grandes números por sua representação decimal. (Na realidade, os números muito grandes não possuem nomes. Por exemplo, como se chamaria o número 10^{1000}?).

Deve ficar claro que o conjunto $\mathbb{N} = \{1, 2, 3, \ldots\}$ dos números naturais é uma seqüência de objetos abstratos que, em princípio, são vazios de significado. Cada um desses objetos (um número natural) possui apenas um lugar determinado nesta seqüência. Nenhuma outra propriedade lhe serve de definição. Todo número tem um sucessor (único) e, com exceção de 1, tem também um único antecessor (número do qual é sucessor).

Vistos desta maneira, podemos dizer que os números naturais são *números ordinais*: 1 é o primeiro, 2 é o segundo, etc.

Um Pequeno Comentário Gramatical

Quando dizemos "o número um", "o número dois" ou "o número três", as palavras "um", "dois" e "três" são substantivos, pois são nomes de objetos. Isto contrasta com o uso destas palavras em frases como "um ano, dois meses e três dias", onde elas aparecem para dar a idéia de número cardinal, isto é, como resultados de contagens. Nesta frase, "um", "dois" e "três" não são substantivos. Pertencem a uma categoria gramatical que, noutras línguas (como francês, inglês e alemão, por exemplo) é chamada *adjetivo numeral* e que os gramáticos brasileiros e portugueses, há um par de décadas, resolveram chamar de *numeral* apenas. Este comentário visa salientar a diferença entre os números naturais, olhados como elementos do conjunto $\mathbb{N}$, e o seu emprego como números cardinais. Este segundo aspecto será abordado no capítulo seguinte.

Recomendação

1. Não se deve dar muita importância à eterna questão de saber se

0 (zero) deve ou não ser incluído entre os números naturais. (Vide "Meu Professor de Matemática", pág. 150.) Praticamente todos os livros de Matemática usados nas escolas brasileiras consideram 0 como o primeiro número natural (conseqüentemente 1 é o segundo, 2 é o terceiro, etc). Como se viu acima, não adotamos esse ponto-de-vista. Trata-se, evidentemente, de uma questão de preferência. Deve-se lembrar que o símbolo 0 (sob diferentes formas gráficas) foi empregado inicialmente pelos maias, posteriormente pelos hindus, difundido pelos árabes e adotado no ocidente, não como um número e sim como um algarismo, com o utilíssimo objetivo de preencher uma casa decimal vazia. (No caso dos maias, a base do sistema de numeração era 20, e não 10.) De resto, a opção do número natural para iniciar a seqüência não se limita a escolher entre 0 e 1. Freqüentemente esquecemos que, do mesmo modo que conhecemos e usamos o zero mas começamos os números naturais com 1, a Matemática grega, segundo apresentada por Euclides, não considerava 1 como um número. Nos "Elementos", encontramos as seguintes definições:

"Unidade é aquilo pelo qual cada objeto é um. Número é uma multitude de unidades".

4. Destaque para o Axioma da Indução

O último dos axiomas de Peano é conhecido como o *axioma da indução*. Ele é a base de um eficiente método de demonstração de proposições referentes a números naturais (demonstrações por indução, ou por recorrência). Enunciado sob a forma de propriedades em vez de conjuntos, ele se formula assim:

Seja $P(n)$ uma propriedade relativa ao número natural n. Suponhamos que

i) $P(1)$ é válida;

ii) Para todo $n \in \mathbb{N}$, a validez de $P(n)$ implica a validez de $P(n')$, onde n' é o sucessor de n.

Então $P(n)$ é válida qualquer que seja o número natural n.

Com efeito, se chamarmos de X o conjunto dos números naturais n para os quais $P(n)$ é válida, veremos que $1 \in X$ em virtude de i) e que $n \in X \Rightarrow n' \in X$ em virtude de ii). Logo, pelo axioma da indução, concluímos que $X = \mathbb{N}$.

Recomendação

2. O axioma da indução é uma forma sagaz e operacional de dizer que qualquer número natural n pode ser alcançado se partirmos de 1 e repetirmos suficientemente a operação de tomar o sucessor de um número. Ele está presente (pelo menos de forma implícita) sempre que, ao afirmarmos a veracidade de uma proposição referente aos números naturais, verificamos que ela é verdadeira para $n = 1$, $n = 2$, $n = 3$ e dizemos "e assim por diante...". Mas é preciso ter cuidado com esta última frase. Ela pressupõe que $P(n) \Rightarrow P(n')$ para todo $n \in \mathbb{N}$. No final deste capítulo, apresentamos como exercicios algumas proposições demonstráveis por recorrência, bem como alguns curiosos paradoxos que resultam do uso inadequado do axioma da indução.

5. Adição e Multiplicação

Entre os números naturais estão definidas duas operações fundamentais: a *adição*, que aos números $n, p \in \mathbb{N}$ faz corresponder a *soma* $n + p$ e a *multiplicação*, que lhes associa o *produto* np.

A soma $n + p$ é o número natural que se obtém a partir de n aplicando-se p vezes seguidas a operação de tomar o sucessor. Em particular, $n + 1$ é o sucessor de n, $n + 2$ é o sucessor do sucessor de n, etc. Por exemplo, tem-se $2 + 2 = 4$ simplesmente porque 4 é o sucessor do sucessor de 2.

De agora em diante, o sucesor do número natural n será designado por $n + 1$.

Quanto ao produto, põe-se $n \cdot 1 = n$ por definição e, quando $p \neq 1$, np é a soma de p parcelas iguais a n.

Em última análise, a soma $n+p$ e o produto np têm mesmo os

significados que lhes são atribuídos pelas explicações dadas acima. Entretanto, até que saibamos utilizar os números naturais para efetuar contagens, não tem sentido falar em "p vezes" e "p parcelas". Por isso, as operações fundamentais devem ser definidas por indução, como se segue.

Adição: $n + 1 =$ sucessor de n e $n + (p + 1) = (n + p) + 1$. Esta última igualdade diz que se sabemos somar p a todos os números naturais n, sabemos também somar $p + 1$: a soma $n + (p + 1)$ é simplesmente o sucessor $(n + p) + 1$ de $n + p$. O axioma da indução garante que a soma $n + p$ está definida para quaisquer $n, p \in \mathbb{N}$.

Multiplicação: $n \cdot 1 = n$ e $n(p + 1) = np + n$. Ou seja: multiplicar um número n por 1 não o altera. E se sabemos multiplicar todos os números naturais n por p, sabemos também multiplicá-los por $p + 1$: basta tomar $n(p + 1) = np + n$. Por indução, sabemos multiplicar todo n por qualquer p.

Estas operações gozam das conhecidas propriedades de associatividade, comutatividade e distributividade. As demonstrações são feitas por indução. (Vide, por exemplo, "Curso de Análise", vol. 1, pág. 28.)

6. Ordem Entre os Números Naturais

Nossa breve descrição do conjunto $\mathbb{N}$ dos números naturais termina com a relação de ordem $m < n$.

Dados $m, n \in \mathbb{N}$, diz-se que m *é menor do que* n, e escreve-se $m < n$, para significar que existe algum $p \in \mathbb{N}$ tal que $n = m + p$. (Isto quer dizer que n é o sucessor do sucessor... do sucessor de m, o ato de tomar o sucessor sendo iterado p vezes.)

A relação $m < n$ tem as seguintes propriedades:

Transitividade: Se $m < n$ e $n < p$ então $m < p$.

Tricotomia: Dados $m, n \in \mathbb{N}$, vale uma, e somente uma, das alternativas: $m = n$, $m < n$ ou $n < m$.

Monotonicidade: Se $m < n$ então, para qualquer $p \in \mathbb{N}$, tem-se $m + p < n + p$ e $mp < np$.

Boa-ordenação: Todo subconjunto não-vazio $X \subset \mathbb{N}$ possui um menor elemento. Isto significa que existe um elemento $m_0 \in X$ que é menor do que todos os demais elementos de X.

A boa-ordenação pode muitas vezes substituir com vantagem a indução como método de prova de resultados referentes a números naturais.

São muito raros e pouco interessantes os exemplos de demonstração por indução que podem ser dados sem usar as operações fundamentais e as desigualdades. Por isso, somente agora apresentamos um deles, seguido de uma demonstração por boa-ordenação.

Exemplo 1. Queremos provar a validez, para todo número natural n, da igualdade

$$P(n): 1 + 3 + 5 + \cdots + (2n - 1) = n^2.$$

Usaremos indução. Para $n = 1$, $P(1)$ se resume a afirmar que $1 = 1$. Supondo $P(n)$ verdadeira para um certo valor de n, somamos $2n + 1$ a ambos os membros da igualdade acima, obtendo

$$1 + 3 + 5 + \cdots + (2n - 1) + (2n + 1) = n^2 + 2n + 1,$$

ou seja:

$$1 + 3 + 5 + \cdots + [2(n + 1) - 1] = (n + 1)^2.$$

Mas esta última igualdade é $P(n + 1)$. Logo $P(n) \Rightarrow P(n + 1)$. Assim, $P(n)$ vale para todo $n \in \mathbb{N}$. Podemos então afirmar que a soma dos n primeiros números ímpares é igual ao quadrado de n.

Exemplo 2. (Usando boa-ordenação.) Lembremos que um número natural p chama-se *primo* quando não pode ser expresso como produto $p = mn$ de dois números naturais, a menos que um deles seja igual a 1 (e o outro igual a p); isto equivale a dizer que os fatores m, n não podem ser ambos menores do que p. Um resultado fundamental em Aritmética diz que todo número natural é primo ou é um produto de fatores primos. Provaremos isto por boa-ordenação. Usaremos a linguagem de conjuntos. Seja X o conjunto

dos números naturais que são primos ou produtos de fatores primos. Observemos que se m e n pertencem a X então o produto mn pertence a X. Seja Y o complementar de X. Assim, Y é o conjunto dos números naturais que não são primos nem são produtos de fatores primos. Queremos provar que Y é vazio. Isto será feito por redução ao absurdo (como sempre se dá nas demonstrações por boa-ordenação). Com efeito, se Y não fosse vazio, haveria um menor elemento $a \in Y$. Então todos os números menores do que a pertenceriam a X. Como a não é primo, ter-se-ia $a = m \cdot n$, com $m < a$ e $n < a$, logo $m \in X$ e $n \in X$. Sendo assim, $mn \in X$. Mas $mn = a$, o que daria $a \in X$, uma contradição. Segue-se que $Y = \varnothing$, concluindo a demonstração.

Exercícios

1. Dado o número natural a, seja $Y \subset \mathbb{N}$ um conjunto com as seguintes propriedades: (1) $a \in Y$; (2) $n \in Y \Rightarrow n+1 \in Y$. Prove que Y contém todos os números naturais maiores do que ou iguais a a. (*Sugestão:* considere o conjunto $X = I_a \cup Y$, onde I_a é o conjunto dos números naturais $\leqslant a$, e prove, por indução, que $X = \mathbb{N}$.)

2. Use o exercício anterior para provar que $2n+1 \leqslant 2^n$ para todo $n \geqslant 2$ e, em seguida, que $n^2 < 2^n$ para todo $n \geqslant 5$.

3. Complete os detalhes da seguinte demonstração do Princípio de Boa Ordenação: Seja $A \subset \mathbb{N}$ um conjunto que não possui um menor elemento. Considere o conjunto X formado pelos números naturais n tais que $1, 2, \ldots, n$ não pertencem a A. Observe que $1 \in X$ e, além disso, se $n \in X$ então todos os elementos de A são $\geqslant n+1$. Como $n+1$ não pode ser o menor elemento de A, conclua que $n+1 \in X$ logo, por indução, segue-se que $X = \mathbb{N}$, portanto A é vazio.

4. Prove, por indução, que

$$\left(\frac{n+1}{n}\right)^n \leqslant n$$

para todo $n \geqslant 3$ e conclua daí que a seqüência

$$1, \sqrt{2}, \sqrt[3]{3}, \sqrt[4]{4} \ldots$$

é decrescente a partir do terceiro termo.

5. Prove, por indução, que

$$1 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

6. Critique a seguinte argumentação: Quer-se provar que todo número natural é pequeno. Evidentemente, 1 é um número pequeno. Além disso, se n for pequeno, $n+1$ também o será, pois não se torna grande um número pequeno simplesmente somando-lhe uma unidade. Logo, por indução, todo número natural é pequeno.

7. Use a distributividade para calcular $(m+n)(1+1)$ de duas maneiras diferentes e em seguida use a lei do corte para concluir que $m + n = n + m$.

8. Seja $X \subset \mathbb{N}$ um conjunto não-vazio, com a seguinte propriedade: para qualquer $n \in \mathbb{N}$, se todos os números naturais menores do que n pertencem a X então $n \in X$. Prove que $X = \mathbb{N}$. (*Sugestão:* boa ordenação.)

9. Seja $P(n)$ uma propriedade relativa ao número natural n. Suponha que $P(1)$, $P(2)$ são verdadeiras e que, para qualquer $n \in \mathbb{N}$, a verdade de $P(n)$ e $P(n+1)$ implica a verdade de $P(n+2)$. Prove que $P(n)$ é verdadeira para todo $n \in \mathbb{N}$.

10. Use indução para provar que

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{1}{4}n^2(n+1)^2.$$

Capítulo 3

Números Cardinais

A importância dos números naturais provém do fato de que eles constituem o modelo matemático que torna possível o processo de contagem. Noutras palavras, eles respondem a perguntas do tipo: "Quantos elementos tem este conjunto?"

Para contar os elementos de um conjunto é necessário usar a noção de correspondência biunívoca, ou bijeção. Trata-se de um caso particular do conceito de função, que abordaremos de forma breve agora e com mais vagar posteriormente.

1. Funções

Dados os conjuntos X, Y, uma *função* $f\colon X \to Y$ (lê-se "uma função de X em Y") é uma regra (ou conjunto de instruções) que diz como associar a cada elemento $x \in X$ um elemento $y = f(x) \in Y$. O conjunto X chama-se o *domínio* e Y é o *contra-domínio* da função f. Para cada $x \in X$, o elemento $f(x) \in Y$ chama-se a *imagem* de x pela função f, ou o *valor* assumido pela função f no ponto $x \in X$. Escreve-se $x \mapsto f(x)$ para indicar que f transforma (ou leva) x em $f(x)$.

Exemplos particularmente simples de funções são a *função identidade* $f\colon X \to X$, definida por $f(x) = x$ para todo $x \in X$ e as *funções constantes* $f\colon X \to Y$, onde se toma um elemento $c \in Y$ e se põe $f(x) = c$ para todo $x \in X$.

Recomendações

1. É importante ressaltar que $f(x)$ é a imagem do elemento $x \in X$

pela função f, ou o valor da função f no ponto $x \in X$. Os livros antigos, bem como alguns atuais, principalmente os de Cálculo, costumam dizer "a função $f(x)$" quando deveriam dizer "a função f". Algumas vezes essa linguagem inexata torna a comunicação mais rápida e fica difícil resistir à tentação de usá-la. Mas é indispensável a cada momento ter a noção precisa do que se está fazendo.

Na prática, há algumas funções com as quais é simples e natural lidar usando a terminologia correta. Por exemplo, é fácil acostumar-se a escrever as funções $\text{sen}: \mathbb{R} \to \mathbb{R}$ e $\log: \mathbb{R}^+ \to \mathbb{R}$, guardando as notações $\text{sen}\, x$ e $\log x$ para os números reais que são os valores destas funções num dado ponto x. Por outro lado, quando se trata de uma função polinomial, o bom-senso nos leva a dizer

$$\text{"a função } x^2 - 5x + 6\text{"}$$

em vez da forma mais correta e mais pedante "a função $p: \mathbb{R} \to \mathbb{R}$ tal que

$$p(x) = x^2 - 5x + 6$$

para todo $x \in \mathbb{R}$". Caso análogo se dá com a função exponencial e^x, embora recentemente se tenha tornado cada vez mais freqüente escrever $\exp(x) = e^x$ e assim poder falar da função $\exp: \mathbb{R} \to \mathbb{R}$.

2. Deve-se ainda observar que uma função consta de três ingredientes: domínio, contra-domínio e a lei de correspondência $x \mapsto f(x)$. Mesmo quando dizemos simplesmente "a função f", ficam subentendidos seu domínio X e seu contra-domínio Y. Sem que eles sejam especificados, não existe a função. Assim sendo, uma pergunta do tipo "Qual é o domínio da função $f(x) = 1/x$"?, estritamente falando, não faz sentido. A pergunta correta seria: "Qual é o maior subconjunto $X \subset \mathbb{R}$ tal que a fórmula $f(x) = 1/x$ define uma função $f: X \to \mathbb{R}$?" Novamente, a pergunta incorreta é mais simples de formular. Se for feita assim, é preciso saber seu significado.

Segue-se do que foi dito acima que as funções $f: X \to Y$ e $g: X' \to$

Y' são iguais se, e somente se, $X = X'$, $Y = Y'$ e $f(x) = g(x)$ para todo $x \in X$.

Exemplos

1. Sejam X o conjunto dos triângulos do plano Π e $\mathbb{R}$ o conjunto dos números reais (que abordaremos logo mais). Se, a cada $t \in X$, fizermos corresponder o número real $f(t) =$ área do triângulo t, obteremos uma função $f: X \to \mathbb{R}$.

2. Sejam S o conjunto dos segmentos de reta do plano Π e Δ o conjunto das retas desse mesmo plano. A regra que associa a cada segmento $AB \in S$ sua mediatriz $g(AB)$ define uma função $g: S \to \Delta$.

3. A correspondência que associa a cada número natural n seu sucessor $n + 1$ define uma função $s: \mathbb{N} \to \mathbb{N}$, com $s(n) = n + 1$.

Uma função $f: X \to Y$ chama-se *injetiva* quando elementos diferentes em X são transformados por f em elementos diferentes em Y. Ou seja, f é injetiva quando

$$x \neq x' \text{ em } X \Rightarrow f(x) \neq f(x').$$

Esta condição pode também ser expressa em sua forma contrapositiva:

$$f(x) = f(x') \Rightarrow x = x'.$$

Nos três exemplos dados acima, apenas o terceiro é de uma função injetiva. (Dois triângulos diferentes podem ter a mesma área e dois segmentos distintos podem ter a mesma mediatriz mas números naturais diferentes têm sucessores diferentes.)

Diz-se que uma função $f: X \to Y$ é *sobrejetiva* quando, para qualquer elemento $y \in Y$, pode-se encontrar (pelo menos) um elemento $x \in X$ tal que $f(x) = y$.

Nos três exemplos dados acima, apenas o segundo apresenta uma função sobrejetiva. (Toda reta do plano é mediatriz de algum segmento mas apenas os números reais positivos podem ser áreas de triângulos e o número 1 não é sucessor de número natural

algum.)

Mais geralmente, chama-se *imagem* do subconjunto $A \subset X$ pela função $f: X \to Y$ ao subconjunto $f(A) \subset Y$ formado pelos elementos $f(x)$, com $x \in A$. A função $f: X \to Y$ é sobrejetiva quando $f(X) = Y$. O conjunto $f(X)$, imagem do domínio X pela função f, chama-se também *a imagem da função* f.

Nos exemplos 1), 2) e 3) a imagem da função f é o conjunto dos números reais positivos, a imagem de g é todo o conjunto Δ e a imagem de s é o conjunto dos números naturais $\geqslant 2$.

Dada a função $f: X \to Y$, para saber se um certo elemento $b \in Y$ pertence ou não à imagem $f(X)$, escrevemos a "equação" $f(x) = b$ e procuramos achar algum $x \in X$ que a satisfaça. Conseqüentemente, para mostrar que f é sobrejetiva deve-se provar que a equação $f(x) = y$ possui uma solução $x \in X$, seja qual for o $y \in Y$ dado.

Recomendação

3. Em muitos exemplos de funções $f: X \to Y$, principalmente na Matemática Elementar, X e Y são conjuntos numéricos e a regra $x \mapsto f(x)$ exprime o valor $f(x)$ por meio de uma fórmula que envolve x. Mas em geral não precisa ser assim. A natureza da regra que ensina como obter $f(x)$ quando é dado x é inteiramente arbitrária, sendo sujeita apenas a duas condições:

a) Não deve haver exceções: a fim de que a função f tenha o conjunto X como domínio, a regra deve fornecer $f(x)$, *seja qual for* $x \in X$ *dado.*

b) Não pode haver ambigüidades: a cada $x \in X$, a regra deve fazer corresponder um *único* $f(x)$ em Y.

Os exemplos a seguir ilustram essas exigências.

Exemplos

4. Considere a tentativa de definir uma função $f: \mathbb{N} \to \mathbb{N}$, estipulando que, para todo $n \in \mathbb{N}$, o número natural

$p = f(x)$ deve ser tal que $p^2 + 3 = n$. O número $p = f(n)$ só pode ser encontrado se n for igual a 4, 7, 12, 19, ... pois nem todos os números naturais são da forma $p^2 + 3$. Assim, esta regra não define uma função com domínio $\mathbb{N}$, porque tem exceções.

5. Indiquemos com X o conjunto dos números reais positivos e com Y o conjunto dos triângulos do plano. Para cada $x \in X$, ponhamos $f(x) = t$ caso t seja um triângulo cuja área é x. Esta regra não define uma função $f: X \to Y$ porque é ambígua: dado o número $x > 0$, existe uma infinidade de triângulos diferentes com área x.

2. A Noção de Número Cardinal

Uma função $f: X \to Y$ chama-se uma *bijeção*, ou uma *correspondência biunívoca* entre X e Y quando é ao mesmo tempo injetiva e sobrejetiva.

Exemplos

6. Sejam $X = \{1, 2, 3, 4, 5\}$ e $Y = \{2, 4, 6, 8, 10\}$. Definindo $f: X \to Y$ pela regra $f(n) = 2n$, temos uma correspondência biunívoca, onde $f(1) = 2$, $f(2) = 4$, $f(3) = 6$, $f(4) = 8$ e $f(5) = 10$.

7. Um exemplo particularmente curioso de correspondência biunívoca foi descoberto pelo físico Galileu Galilei, que viveu há quatrocentos anos. Seja P o conjunto dos números naturais pares: $P = \{2, 4, 6, \ldots, 2n, \ldots\}$. Obtém-se uma correspondência biunívoca $f: \mathbb{N} \to P$ pondo-se $f(n) = 2n$ para todo $n \in \mathbb{N}$. O interessante deste exemplo é que P é um subconjunto próprio de $\mathbb{N}$.

8. Sejam Y a base de um triângulo e X um segmento paralelo a Y, unindo os outros dois lados desse triângulo. Seja ainda P o vértice oposto à base Y. Obtém-se uma correspondência biunívoca $f: X \to Y$ associando a cada $x \in X$ o ponto $f(x)$ onde a semi-reta Px intersecta a base Y

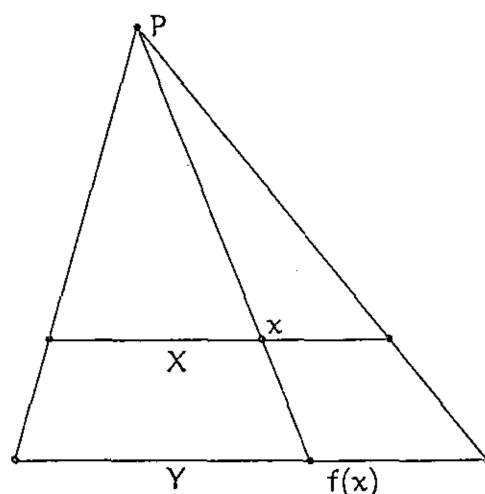


Figura 2

9. Neste exemplo, $X = C - \{P\}$ é o conjunto obtido retirando da circunferência o ponto P e Y é uma reta perpendicular ao diâmetro que passa por P.

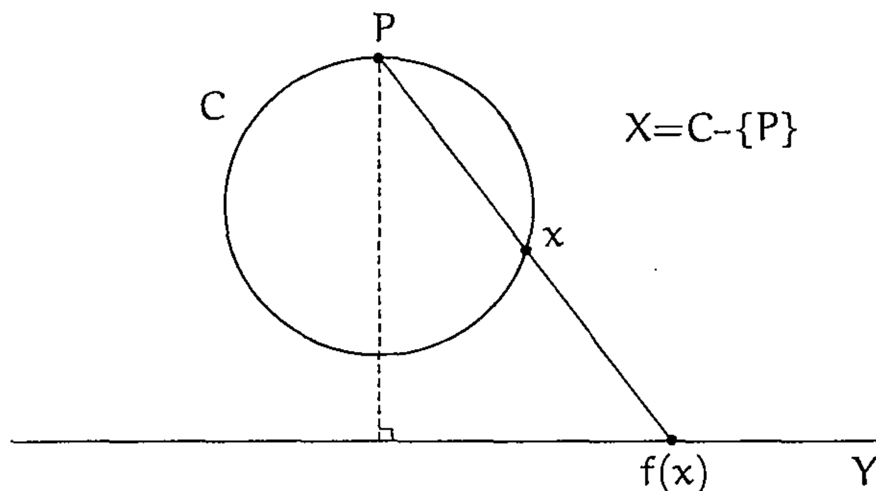


Figura 3

Definiremos uma correspondência biunívoca $f: X \to Y$ pondo, para cada $x \in X$, $f(x)$ = interseção da semi-reta Px com a reta Y.

Diz-se que dois conjuntos X e Y têm o *mesmo número cardinal* quando se pode definir uma correspondência biunívoca $f: X \to Y$.

Cada um dos quatro exemplos acima exibe um par de conjuntos X, Y com o mesmo número cardinal.

Exemplo 10. Sejam $X = \{1\}$ e $Y = \{1, 2\}$. Evidentemente não

pode existir uma correspondência biunívoca $f: X \rightarrow Y$, portanto X e Y não têm o mesmo número cardinal.

A palavra "número" no dicionário

As vezes se diz que os conjuntos X e Y são (numericamente) *equivalentes* quando é possível estabelecer uma correspondência biunívoca $f: X \rightarrow Y$, ou seja, quando X e Y têm o mesmo número cardinal.

Isto explica (embora não justifique) a definição dada no dicionário mais vendido do país. Em algumas situações, ocorrem em Matemática definições do tipo seguinte: um *vetor* é o conjunto de todos os segmentos de reta do plano que são equipolentes a um segmento dado. (Definição "por abstração".) Nessa mesma veia, poder-se-ia tentar dizer: "número cardinal de um conjunto é o conjunto de todos os conjuntos equivalentes a esse conjunto." No caso do dicionário, há um conjunto de defeitos naquela definição, com um número cardinal razoavelmente elevado. Os três mais graves são:

1. Um dicionário não é um compêndio de Matemática, e muito menos de Lógica. Deve conter explicações acessíveis ao leigo (de preferência, corretas). As primeiras acepções da palavra "número" num dicionário deveriam ser "quantidade" e "resultado de uma contagem ou de uma medida".

2. A definição em causa só se aplica a números cardinais, mas a idéia de número deveria abranger os racionais e, pelo menos, os reais.

3. O "conjunto de todos os conjuntos equivalentes a um conjunto dado" é um conceito matematicamente incorreto. A noção de conjunto não pode ser usada indiscriminadamente, sem submeter-se a regras determinadas, sob pena de conduzir a paradoxos, ou contradições. Uma dessas regras proibe que se forme conjuntos a não ser que seus elementos pertençam a, ou sejam subconjuntos de, um determinado conjunto-universo. Um exemplo de paradoxo que resulta da desatenção a essa regra é "o conjunto X de todos os conjuntos que não são elementos de si mesmos." Pergunta-se: X é

ou não é um elemento de si mesmo? Qualquer que seja a resposta, chega-se a uma contradição.

3. Conjuntos Finitos

Dado $n \in \mathbb{N}$, indiquemos com a notação I_n o conjunto dos números naturais de 1 até n. Assim, $I_1 = \{1\}$, $I_2 = \{1,2\}$, $I_3 = \{1,2,3\}$ e, mais geralmente, um número natural k pertence a I_n se, e somente se, $1 \leqslant k \leqslant n$.

Seja X um conjunto. Diz-se que X é *finito*, e que X *tem* n *elementos* quando se pode estabelecer uma correspondência biunívoca $f: I_n \to X$. O número natural n chama-se então o *número cardinal* do conjunto X ou, simplesmente, o número de elementos de X. A correspondência $f: I_n \to X$ chama-se uma *contagem* dos elementos de X. Pondo $f(1) = x_1$, $f(2) = x_2, \ldots, f(n) = x_n$, podemos escrever $X = \{x_1, x_2, \ldots, x_n\}$. Para todo n, o conjunto I_n é finito e seu número cardinal é n. Assim, todo número natural n é o número cardinal de algum conjunto finito.

A fim de evitar exceções, admite-se ainda incluir o conjunto vazio $\varnothing$ entre os conjuntos finitos e diz-se que $\varnothing$ tem zero elementos. Assim, por definição, zero é o número cardinal do conjunto vazio.

Diz-se que um conjunto X é *infinito* quando ele não é finito. Isto quer dizer que X não é vazio e que, não importa qual seja $n \in \mathbb{N}$, não existe correspondência biunívoca $f: I_n \to X$.

No Exemplo 6 acima, temos $X = I_5$ e $f: X \to Y$ é uma contagem dos elementos de Y. Assim, Y é um conjunto finito, com 5 elementos. O conjunto $\mathbb{N}$ dos números naturais é infinito. Com efeito, dada qualquer função $f: I_n \to \mathbb{N}$, não importa qual n se fixou, pomos $k = f(1) + f(2) + \cdots + f(n)$ e vemos que, para todo $x \in I_n$, tem-se $f(x) < k$, logo não existe $x \in I_n$ tal que $f(x) = k$. Assim, é impossível cumprir a condição b) da definição de correspondência biunívoca.

O número cardinal de um conjunto finito X, que indicaremos com a notação $n(X)$, goza de algumas propriedades básicas, entre

as quais destacaremos as seguintes:

1. *O número de elementos de um conjunto finito é o mesmo, seja qual for a contagem que se adote.*

Isto significa que se $f: I_m \to X$ e $g: I_n \to X$ são correspondências biunívocas então $m = n$.

2. *Todo subconjunto* Y *de um conjunto finito* X *é finito e* $n(Y) \leqslant n(X)$. *Tem-se* $n(Y) = n(X)$ *somente quando* $Y = X$.

3. *Se* X *e* Y *são finitos então* $X \cup Y$ *é finito e tem-se* $n(X \cup Y) = n(X) + n(Y) - n(X \cap Y)$.

4. *Sejam* X, Y *conjuntos finitos. Se* $n(X) > n(Y)$, *nenhuma função* $f: X \to Y$ *é injetiva e nenhuma função* $g: Y \to X$ *é sobrejetiva.*

As demonstrações destes fatos se fazem por indução ou por boa-ordenação. (Veja, por exemplo, *Curso de Análise*, vol. 1, págs. 33–38.) A primeira parte do item 4. acima é conhecida como o *princípio das casas de pombos*: se há mais pombos do que casas num pombal, qualquer modo de alojar os pombos deverá colocar pelo menos dois deles na mesma casa. Às vezes, o mesmo fato é chamado o *princípio das gavetas*: se $m > n$, qualquer maneira de distribuir m objetos em n gavetas deverá pôr ao menos dois desses objetos na mesma gaveta. (Na referência citada, este é o Corolário 1 na página 35.)

O princípio das casas de pombos, com toda sua simplicidade, possui interessantes aplicações. Vejamos duas delas.

Exemplo 13. Tomemos um número natural de 1 a 9. Para fixar as idéias, seja 3 esse número. Vamos provar que todo número natural m possui um múltiplo cuja representação decimal contém apenas os algarismos 3 ou 0. Para isso, consideremos o conjunto $X = \{3, 33, \ldots, 33\ldots3\}$, cujos elementos são os m primeiros números naturais representados somente por algarismos iguais a 3. Se algum dos elementos de X for múltiplo de m, nosso trabalho acabou. Caso contrário, formamos o conjunto $Y = \{1, 2, \ldots, m-1\}$ e definimos a função $f: X \to Y$ pondo, para cada $x \in X$,

$$f(x) = \text{ resto da divisão de } x \text{ por } m.$$

Como X tem mais elementos do que Y, o princípio das casas de pombos assegura que existem elementos $x_1 < x_2$ no conjunto X tais que $f(x_1) = f(x_2)$. Isto significa que x_1 e x_2, quando divididos por m, deixam o mesmo resto. Logo $x_2 - x_1$ é múltiplo de m. Mas é claro que se x_1 tem p algarismos e x_2 tem $p + q$ algarismos então a representação decimal de $x_2 - x_1$ consiste em q algarismos iguais a 3 seguidos de p algarismos iguais a 0.

Exemplo 14. Vamos usar o princípio das gavetas para provar que, numa reunião com n pessoas ($n \geqslant 2$), há sempre duas pessoas (pelo menos) que têm o mesmo número de amigos naquele grupo. Para ver isto, imaginemos n caixas, numeradas com $0, 1, \ldots, n-1$. A cada uma das n pessoas entregamos um cartão que pedimos para depositar na caixa correspondente ao número de amigos que ela tem naquele grupo. As caixas de números 0 e $n - 1$ não podem ambas receber cartões pois se houver alguém que não tem amigos ali, nenhum dos presentes pode ser amigo de todos, e vice-versa. Portanto temos, na realidade, n cartões para serem depositados em $n - 1$ caixas. Pelo princípio das gavetas, pelo menos uma das caixas vai receber dois ou mais cartões. Isto significa que duas ou mais pessoas ali têm o mesmo número de amigos entre os presentes.

4. Sobre Conjuntos Infinitos

Para encerrar estas considerações a respeito de números cardinais, faremos alguns comentários sobre conjuntos infinitos.

Em primeiro lugar, convém esclarecer que a maior contribuição de Cantor não foi a adoção da linguagem e da notação dos conjuntos e sim suas descobertas sobre os números cardinais de conjuntos infinitos. Ele foi o primeiro a descobrir que existem conjuntos infinitos com diferentes cardinalidades ao provar que não pode haver uma correspondência biunívoca entre $\mathbb{N}$ e o conjunto $\mathbb{R}$ dos números reais e que nenhum conjunto X pode estar em correspondência biunívoca com o conjunto $\mathcal{P}(X)$ cujos elementos são os subconjuntos de X. Além disso, ele mostrou que a reta, o

plano e o espaço tri-dimensional (ou mesmo espaços com dimensão superior a três) têm o mesmo número cardinal. Estes fatos, que atualmente são considerados corriqueiros entre os matemáticos, causaram forte impacto na época (meados do século dezenove).

A segunda observação diz respeito a funções $f: X \to X$ de um conjunto em si mesmo. Quando X é finito, f é injetiva se, e somente se, é sobrejetiva. (Vide referência anterior.) Mas isto não é verdadeiro para X infinito. Por exemplo, se definirmos a função $f: \mathbb{N} \to \mathbb{N}$ pondo, para cada $n \in \mathbb{N}$, $f(n)$ = número de fatores primos distintos que ocorrem na decomposição de n, veremos que f é sobrejetiva mas não é injetiva. (Para cada $b \in \mathbb{N}$ existe uma infinidade de números n tais que $f(n) = b$.) Além disso, as funções $f: \mathbb{N} \to \mathbb{N}$, $g: \mathbb{N} \to \mathbb{N}$, $h: \mathbb{N} \to \mathbb{N}$ e $\varphi: \mathbb{N} \to \mathbb{N}$, definidas por

$$f(n) = n + 1,$$

$$g(n) = n + 30,$$

$$h(n) = 2n \text{ e}$$

$$\varphi(n) = 3n$$

são injetivas mas não são sobrejetivas. Estas quatro funções são protagonistas da seguinte historinha que fecha a seção.

Fantasia Matemática

O Grande Hotel Georg Cantor tinha uma infinidade de quartos, numerados consecutivamente, um para cada número natural. Todos eram igualmente confortáveis. Num fim-de-semana prolongado, o hotel estava com seus quartos todos ocupados, quando chega um viajante. A recepcionista vai logo dizendo:

– Sinto muito, mas não há vagas.

Ouvindo isto, o gerente interveio:

– Podemos abrigar o cavalheiro, sim senhora.

E ordena:

– Transfira o hóspede do quarto 1 para o quarto 2, passe o do quarto 2 para o quarto 3 e assim por diante. Quem estiver no

quarto n, mude para o quarto $n + 1$. Isto manterá todos alojados e deixará disponível o quarto 1 para o recém-chegado.

Logo depois chegou um ônibus com 30 passageiros, todos querendo hospedagem. A recepcionista, tendo aprendido a lição, removeu o hóspede de cada quarto n para o quarto $n + 30$ e acolheu assim todos os passageiros do ônibus. Mas ficou sem saber o que fazer quando, horas depois, chegou um trem com uma infinidade de passageiros. Desesperada, apelou para o gerente que prontamente resolveu o problema dizendo: – Passe cada hóspede do quarto n para o quarto $2n$. Isto deixará vagos todos os apartamentos de número ímpar, nos quais poremos os novos hóspedes. – Pensando melhor: mude quem está no quarto n para o quarto $3n$. Os novos hóspedes, ponha-os nos quartos de número $3n + 2$. Deixaremos vagos os quartos de número $3n + 1$. Assim, sobrarão ainda infinitos quartos vazios e eu poderei ter sossego por algum tempo.

Recomendação

4. Não confunda conjunto infinito com aquele que tem um número muito grande (porém finito) de elementos. Quando, na linguagem comum, se diz algo como "– Já ouvi isto uma infinidade de vezes", trata-se de uma mera força de expressão. Não há distâncias infinitas (mesmo entre duas galáxias bem afastadas) e até o número de átomos do universo é finito. (O físico Arthur Eddington estimou o número de prótons do universo em 136×2^{256}. O número de átomos é certamente menor pois todo átomo contém ao menos um próton.) É importante ter sempre em mente que nenhum número natural n é maior do que todos os demais: tem-se sempre $n < n + 1$.

Exercícios

1. Seja $f: X \to Y$ uma função. A *imagem inversa* por f de um conjunto $B \subset Y$ é o conjunto $f^{-1}(B) = \{x \in x; f(x) \in B\}$. Prove que

se tem sempre $f^{-1}(f(A)) \supset A$ para todo $A \subset X$ e $f(f^{-1}(B)) \subset B$ para todo $B \subset Y$. Prove também que f é injetiva se, e somente se, $f^{-1}(f(A)) = A$ para todo $A \subset X$. Analogamente, mostre que f é sobrejetiva se, e somente se, $f(f^{-1}(B)) = B$ para todo $B \subset Y$.

2. Prove que a função $f: X \to Y$ é injetiva se, e somente se, existe uma função $g: Y \to X$ tal que $g(f(x)) = x$ para todo $x \in X$.

3. Prove que a função $f: X \to Y$ é sobrejetiva se, e somente se, existe uma função $h: Y \to X$ tal que $f(h(y)) = y$ para todo $y \in Y$.

4. Dada a função $f: X \to Y$, suponha que $g, h: Y \to X$ são funções tais que $g(f(x)) = x$ para todo $x \in X$ e $f(h(y)) = y$ para todo $y \in Y$. Prove que $g = h$.

5. Defina uma função sobrejetiva $f: \mathbb{N} \to \mathbb{N}$ tal que, para todo $n \in \mathbb{N}$, a equação $f(x) = n$ possui uma infinidade de raízes $x \in \mathbb{N}$. (*Sugestão:* todo número natural se escreve, de modo único sob a forma $2^a \cdot b$, onde $a, b \in \mathbb{N}$ e b é ímpar.)

6. Prove, por indução, que se X é um conjunto finito com n elementos então existem $n!$ bijeções $f: X \to X$.

7. Qual o erro da seguinte demonstração por indução:

Teorema: *Todas as pessoas têm a mesma idade.*

Prova: Provaremos por indução que se X é um conjunto de n ($n \geqslant 1$) pessoas, então todos os elementos de X têm a mesma idade.

Se $n = 1$ a afirmação é evidentemente verdadeira pois se X é um conjunto formado por uma única pessoa, todos os elementos de X têm a mesma idade.

Suponhamos agora que a afirmação seja verdadeira para todos os conjuntos de n elementos. Consideremos um conjunto com $n + 1$ pessoas, $\{a_1, a_2, \ldots, a_n, a_{n+1}\}$. Ora, $\{a_1, a_2, \ldots, a_n\}$ é um conjunto de n pessoas, logo $a_1, a_2, \ldots, a_n$ têm a mesma idade. Mas $\{a_2, \ldots, a_n, a_{n+1}\}$ ta,bém é um conjunto de n elementos, logo todos os seus elementos, em particular a_n e a_{n+1}, têm a mesma idade. Mas de $a_1, a_2, \ldots, a_n$ têm a mesma idade e a_n e a_{n+1} têm a mesma

idade, todos os elementos de $\{a_1, a_2, \ldots, a_n, a_{n+1}\}$ têm a mesma idade, conforme queríamos demonstrar.

8. Prove, por indução, que um conjunto com n elementos possui 2^n subconjuntos.

9. Dados n ($n \geqslant 2$) objetos de pesos distintos, prove que é possível determinar qual o mais leve e qual o mais pesado fazendo $2n - 3$ pesagens em uma balança de pratos. É esse o número mínimo de pesagens que permitem determinar o mais leve e o mais pesado?

10. Prove que, dado um conjunto com n elementos, é possível fazer uma fila com seus subconjuntos de tal modo que cada subconjunto da fila pode ser obtido a partir do anterior pelo acréscimo ou pela supressão de um único elemento.

11. Todos os quartos do Hotel Georg Cantor estão ocupados, quando chegam os trens $T_1, T_2, \ldots, T_n, \ldots$ (em quantidade infinita), cada um deles com infinitos passageiros. Que deve fazer o gerente para hospedar todos?

Capítulo 4

Números Reais

Nos dois capítulos anteriores, foram introduzidos os números naturais e foi mostrado como eles são empregados na operação de contagem. Veremos agora de que modo o processo de medição das grandezas ditas contínuas conduz à noção de número real. Usaremos como protótipo a determinação do comprimento de um segmento de reta. Este exemplo de medição é tão significativo que o conjunto dos números reais é também conhecido como a *reta real* ou, simplesmente, a *reta*.

1. Segmentos Comensuráveis e Incomensuráveis

Seja AB um segmento de reta. Para medi-lo, é necessário fixar um segmento-padrão u, chamado *segmento unitário*. Por definição, a medida do segmento u é igual a 1. Estipularemos ainda que segmentos congruentes tenham a mesma medida e que se $n-1$ pontos interiores decompuserem AB em n segmentos justapostos então a medida de AB será igual à soma das medidas desses n segmentos. Se estes segmentos parciais forem todos congruentes a u, diremos que u *cabe* n *vezes em* AB e a medida de AB (que representaremos por $\overline{AB}$) será igual a n.

Pode ocorrer que o segmento unitário não caiba um número exato de vezes em AB. Então a medida de AB não será um número natural. Esta situação conduz à idéia de *fração*, conforme mostraremos agora.

Procuramos um pequeno segmento de reta w, que caiba n vezes no segmento unitário u e m vezes em AB. Este segmento w será então uma *medida comum* de u e AB. Encontrado w, diremos

que AB e u são *comensuráveis*. A medida de w será a fração $1/n$ e a medida de AB, por conseguinte, será m vezes $1/n$, ou seja, igual a m/n.

Relutantes em admitir como número qualquer objeto que não pertencesse ao conjunto $\{2, 3, 4, 5, \ldots\}$, os matemáticos gregos à época de Euclides não olhavam para a fração m/n como um número e sim como uma razão entre dois números, igual à razão entre os segmentos AB e u.

Na realidade, não é muito importante que eles chamassem m/n de número ou não, desde que soubessem, como sabiam, raciocinar com esses símbolos. (Muito pior eram os egípcios que, com exceção de 2/3, só admitiam frações de numerador 1. Todas as demais, tinham que ser expressas como somas de frações de numerador 1 e denominadores diferentes. Por exemplo, 7/10 no Egito era escrito como $1/3 + 1/5 + 1/6$.)

O problema mais sério é que por muito tempo se pensava que dois segmentos quaisquer eram sempre comensuráveis: sejam quais fossem AB e CD, aceitava-se tacitamente que haveria sempre um segmento EF que caberia um número exato n de vezes em AB e um número exato m de vezes em CD. Esta crença talvez adviesse da Aritmética, onde dois números naturais quaisquer têm sempre um divisor comum (na pior hipótese, igual a 1).

A ilusão da comensurabilidade durou até o quarto século antes de Cristo. Naquela época, em Crotona, sul da Itália, havia uma seita filosófico-religiosa, liderada por Pitágoras. Um dos pontos fundamentais de sua doutrina era o lema "Os números governam o mundo". (Lembremos que números para eles eram números naturais, admitindo-se tomar razões entre esses números, formando as frações.) Uma enorme crise, que abalou os alicerces do pitagorismo e, por algum tempo, toda a estrutura da Matemática grega, surgiu quando, entre os próprios discípulos de Pitágoras, alguém observou que o lado e a diagonal de um quadrado são segmentos de reta incomensuráveis.

O argumento é muito simples e bem conhecido.

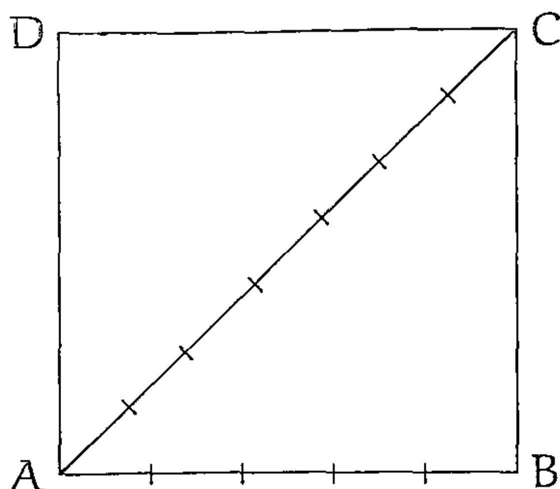


Figura 4

Se houvese um segmento de reta u que coubesse n vezes no lado AB e m vezes na diagonal AC do quadrado ABCD então, tomando AB como unidade de comprimento, a medida de AC seria igual a m/n enquanto, naturalmente, a medida de AB seria 1. Pelo Teorema de Pitágoras teríamos $(m/n)^2 = 1^2 + 1^2$, donde $m^2/n^2 = 2$ e $m^2 = 2n^2$. Mas esta última igualdade é absurda, pois na decomposição de m^2 em fatores primos o expoente do fator 2 é par enquanto em $2n^2$ é ímpar.

A existência de segmentos incomensuráveis significa que os números naturais mais as frações são insuficientes para medir todos os segmentos de reta.

A solução que se impunha, e que foi finalmente adotada, era a de ampliar o conceito de número, introduzindo os chamados *números irracionais*, de tal modo que, fixando uma unidade de comprimento arbitrária, qualquer segmento de reta pudesse ter uma medida numérica. Quando o segmento considerado é comensurável com a unidade escolhida, sua medida é um número *racional* (inteiro ou fracionário). Os números irracionais representam medidas de segmentos que são incomensuráveis com a unidade. No exemplo acima, quando o lado do quadrado mede 1, a medida da diagonal é o número irracional $\sqrt{2}$. (O fato de que esta conclusão não depende do tamanho do quadrado que se considera, deve-se a que dois quadrados quaisquer são figuras semelhantes.)

Recomendação

1. Nos meios de comunicação e entre pessoas com limitado conhecimento matemático, a palavra *incomensurável* é muitas vezes usada em frases do tipo: havia um número incomensurável de formigas em nosso piquenique. Nunca diga isso. Incomensurabilidade é uma relação entre duas grandezas da mesma espécie; não dá idéia de quantidade muito grande. Uma palavra adequada no caso das formigas seria *incontável*. Noutros casos, como um campo gigantesco, poderia ser *imensurável*. Mas nada é incomensurável, a não ser quando comparado com outro objeto (grandeza) da mesma espécie.

2. A Reta Real

A fim de ganhar uma idéia mais viável dos novos números, que denominamos irracionais e, em particular, situá-los em relação aos racionais, imaginamos uma reta, na qual foram fixados um ponto O, chamado a *origem*, e um ponto A, diferente de O. Tomaremos o segmento OA como unidade de comprimento. A reta OA será chamada a *reta real*, ou o *eixo real*.

A origem O divide a reta em duas semi-retas. A que contém A chama-se a semi-reta *positiva*. A outra é a semi-reta *negativa*. Diremos que os pontos da semi-reta positiva estão à direita de O e os da semi-reta negativa à esquerda de O.

Seja X um ponto qualquer da reta OA. Se o segmento de reta OA couber um número exato n de vezes em OX, diremos que a *abcissa* de X é o número natural n ou o número *negativo* $-n$, conforme X esteja à direita ou à esquerda da origem. Se X coincidir com a origem, sua abcissa será 0 (zero).

O conjunto $\mathbb{Z}$, formado pelo número zero e pelas abcissas dos pontos X do eixo real, tais que o segmento unitário cabe um número exato de vezes em OX, chama-se o conjunto dos *números inteiros*. Ele é a reunião $\mathbb{Z} = \mathbb{N} \cup \{0\} \cup (-\mathbb{N})$, dos números naturais com o zero e o conjunto $-\mathbb{N}$ dos números negativos.

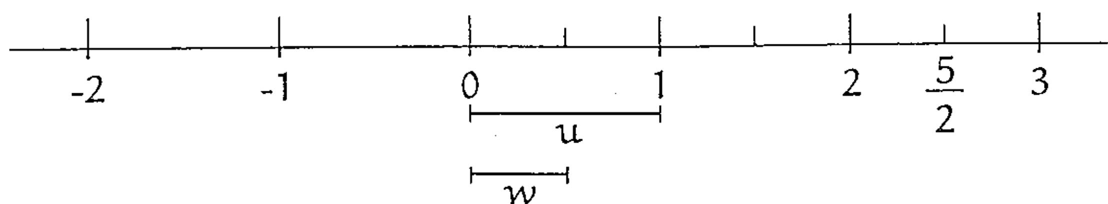


Figura 5

Mais geralmente, se o ponto X, pertencente ao eixo real, é tal que o segmento OX é comensurável com o segmento unitário OA, de modo que algum segmento w caiba n vezes em OA e m vezes em OX, diremos que a abcissa do ponto X é m/n ou $-m/n$, conforme X esteja à direita ou à esquerda da origem. Isto inclui, naturalmente, o caso em que o segmento OA cabe um número exato de vezes em OX, quando se tem $n = 1$ e a abcissa de X pertence a $\mathbb{Z}$.

O conjunto $\mathbb{Q}$, formado pelas abcissas dos pontos X do eixo real tais que o segmento OX é comensurável com o segmento unitário A chama-se o conjunto dos *números racionais*. Tem-se $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$. Como vimos acima, os números racionais são representados por frações m/n, onde $m \in \mathbb{Z}$ e $n \in \mathbb{N}$.

Se, agora, tomarmos um ponto X no eixo real de tal modo que os segmentos OX e OA sejam incomensuráveis, inventaremos um número x, que chamaremos de *número irracional*, e diremos que x é a abcissa do ponto X. O número x será considerado positivo ou negativo, conforme o ponto X esteja à direita ou à esquerda da origem, respectivamente. Quando X está à direita da origem, x é, por definição, a medida do segmento OX. Se X está à esquerda da origem, a abcissa x é essa medida precedida do sinal menos.

O conjunto $\mathbb{R}$, cujos elementos são os números racionais e os números irracionais chama-se o conjunto dos *números reais*. Existe uma correspondência biunívoca entre a reta OA e o conjunto $\mathbb{R}$, a qual associa a cada ponto X dessa reta sua abcissa, isto é, a medida do segmento OX, ou esta medida precedida do sinal menos.

Temos $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.

Observação. As letras $\mathbb{N}$, $\mathbb{Q}$ e $\mathbb{R}$ são as iniciais das palavras

número (ou *natural*), *quociente* e *real*. A letra $\mathbb{Z}$ é a inicial da palavra *zahl*, que significa número em alemão.

O conjunto $\mathbb{R}$ pode ser visto como o modelo aritmético de uma reta enquanto esta, por sua vez, é o modelo geométrico de $\mathbb{R}$. Esta inter-relação entre Geometria e Aritmética, entre pontos e números, é responsável por grandes progressos da Matemática atual.

A interpretação dos números reais como abcissas dos pontos de uma reta fornece uma visão intuitiva bastante esclarecedora sobre a soma $x + y$ e a relação de ordem $x < y$, com $x, y \in \mathbb{R}$.

Com efeito, se X e Y são os pontos dos quais x e y respectivamente são as abcissas, diz-se que x é *menor* do que y, e escreve-se $x < y$ quando X está à esquerda de Y, isto é, quando o sentido de percurso de X para Y é o mesmo de O para A. Quanto à soma, $x + y$ é a abcissa do ponto Y' tal que o segmento XY' tem o mesmo comprimento e o mesmo sentido de percurso de OY.

Também o produto xy dos números reais x, y pode ser definido geometricamente, de acordo com a figura abaixo, quando $x > 0$ e $y > 0$. Nos demais casos, é só mudar o sinal de xy convenientemente.

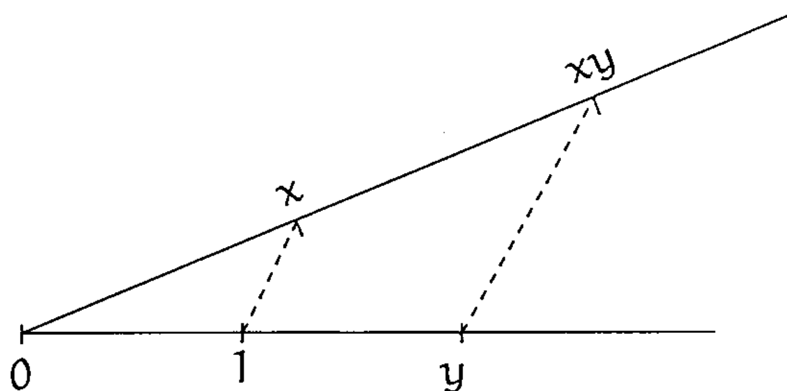


Figura 6: O produto de números reais, visto geometricamente.

As construções geométricas que fornecem interpretações visuais para a soma e para o produto de números reais já eram conhecidas desde Euclides (300 anos antes de Cristo). Vale lembrar apenas que elas representavam operações sobre grandezas

(no caso, segmentos de reta), não sobre números reais.

O progresso da Ciência e a diversidade de aplicações da Matemática, dos casos mais corriqueiros até a alta tecnologia, há muito tempo deixaram claro que esta visão geométrica, por mais importante que tenha sido e ainda seja, precisa ser complementada por uma descrição algébrica de $\mathbb{R}$. Tal complementação requer que seja feita uma lista das propriedades (axiomas) do conjunto $\mathbb{R}$, a partir das quais todos os fatos sobre números reais possam ser demonstrados. Algo parecido com os axiomas de Peano para os números naturais. Só que, naturalmente, uma estrutura mais elaborada, pois $\mathbb{R}$ é uma concepção bem mais rica e mais sutil do que $\mathbb{N}$.

A descrição mais simples de $\mathbb{R}$ consiste em dizer que se trata de um *corpo ordenado completo*. Os detalhes dessa caracterização não são difíceis, mas escapam aos nossos objetivos aqui. O leitor interessado pode consultar o livro *Análise Real*, vol. 1, capítulo 2.

Diremos apenas que $\mathbb{R}$ é um *corpo* porque estão definidas aí as quatro operações: adição, subtração, multiplicação e divisão. É um corpo *ordenado* porque existe a relação $x < y$, que está interligada com a adição e a multiplicação pelas leis conhecidas de monotonicidade. E, finalmente, a *completeza* de $\mathbb{R}$ equivale à continuidade da reta. É ela que garante a existência de $\sqrt[n]{a}$ e, mais geralmente, de a^x para todo $a > 0$ e todo $x \in \mathbb{R}$. É a completeza de $\mathbb{R}$ que diferencia os reais dos racionais pois, afinal de contas, $\mathbb{Q}$ também é um corpo ordenado, só que não é completo. Há várias maneiras de formular matematicamente a afirmação de que o corpo dos números reais é completo. Todas elas envolvem, direta ou indiretamente, a idéia de aproximação, ou limite. Na próxima seção veremos um exemplo de como a completeza de $\mathbb{R}$ se faz necessária para assegurar que toda expressão decimal representa um número real.

Na prática, nossos olhos (e mesmo os instrumentos mais delicados de aferição) têm um extremo de percepção (ou de precisão), sendo incapazes de distinguir diferenças inferiores a esse extremo.

Portanto nenhuma medição experimental pode oferecer como resultado um número irracional. Deve-se entretanto lembrar que, quando o raciocínio matemático assegura a incomensurabilidade, o número racional obtido experimentalmente é apenas um valor aproximado; o valor exato é um número irracional.

3. Expressões Decimais

Para efetuar cálculos, a forma mais eficiente de representar os números reais é por meio de expressões decimais. Vamos falar um pouco sobre elas. É claro que basta considerar os números reais positivos. Para tratar de números negativos, simplesmente se acrescenta o sinal menos.

Uma *expressão decimal* é um símbolo da forma

$$\alpha = a_0{,}a_1a_1\ldots a_n\ldots,$$

onde a_0 é um número inteiro $\geqslant 0$ e $a_1, a_2, \ldots, a_n, \ldots$ são *dígitos*, isto é, números inteiros tais que $0 \leqslant a_n \leqslant 9$. Para cada $n \in \mathbb{N}$, tem-se um dígito a_n, chamado o *n-ésimo dígito* da expressão decimal α. O número natural a_0 chama-se a *parte inteira* de α.

Exemplo 1. $\alpha = 13{,}42800\ldots$, $\beta = 25{,}121212\ldots$ e $\pi = 3{,}1415\,9265\ldots$ são expressões decimais. Nos casos de α e β, está claro como se obtêm os dígitos que não estão explicitados. No caso de π (medida da circunferência quando se toma o diâmetro como unidade), o que está escrito aqui não permite saber qual a regra para achar os dígitos a partir do nono, mas existem processos bem definidos e eficientes para calculá-los. Recentemente, com auxílio de algoritmos especialmente concebidos e computadores rápidos, foi possível determinar os primeiros 56 bilhões de dígitos de π.

Mas de que forma uma seqüência de dígitos, precedida de um número inteiro, representa um número real? A resposta é: a expressão decimal α, dada acima, representa o número real

$$(*) \qquad \overline{\alpha} = a_0 + \frac{a_1}{10} + \frac{a_2}{10^2} + \cdots + \frac{a_n}{10^n} + \cdots.$$

Na realidade, é meio pedante usar uma notação diferente, $\overline{\alpha}$,

para indicar o número real cuja expressão decimal é α. Na prática, não se faz isso. Vamos então seguir o costume e usar a mesma notação α, para o número e sua expressão decimal.

Mais importante é explicar o significado daquelas reticências no final da igualdade. Elas dão a entender de que se trata de uma soma com infinitas parcelas, mas isto é uma coisa que não tem sentido, pelo menos em princípio. O significado da igualdade (*) é o seguinte: o número real α (que já estamos escrevendo sem a barra) tem por valores aproximados os números racionais

$$\alpha_n = a_0 + \frac{a_1}{10} + \cdots + \frac{a_n}{10^n}. \quad (n = 0, 1, 2, \ldots).$$

Quando se substitui α por α_n, o erro cometido não é superior a

$$\frac{1}{10^n} = 10^{-n}.$$

Assim, a_0 é o maior número natural contido em α, a_1 é o maior dígito tal que

$$a_0 + \frac{a_1}{10} \leqslant \alpha,$$

a_2 é o maior dígito tal que

$$a_0 + \frac{a_1}{10} + \frac{a_2}{10^2} \leqslant \alpha, \quad \text{etc.}$$

Deste modo, tem-se uma seqüência não-decrescente de números racionais

$$\alpha_0 \leqslant \alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_n \leqslant \cdots$$

que são valores (cada vez mais) aproximados do número real α. Mais precisamente, tem-se $0 \leqslant \alpha - \alpha_n \leqslant 10^{-n}$ para cada $n = 0, 1, 2, 3, 4 \ldots$.

Diz-se então que o número real α é o *limite* desta seqüência de números racionais. O fato de que *existe* sempre um número real que é limite desta seqüência (isto é, que tem os α_n como seus valores aproximados) é uma forma de dizer que o corpo ordenado dos números reais é completo.

Há algumas situações particulares que merecem ser vistas separadamente.

A primeira é quando, a partir de um certo ponto, todos os dígitos a_n se tornam iguais a zero:

$$\alpha = a_0, a_1 a_2 \ldots a_n 0 0 0 \ldots$$

Então

$$\alpha = a_0 + \frac{a_1}{10} + \cdots + \frac{a_n}{10^n}$$

é um número racional; na realidade uma fração decimal (fração cujo denominador é uma potência de 10). Por exemplo

$$13,42800\ldots = 13 + \frac{4}{10} + \frac{1}{100} + \frac{8}{1000} = \frac{13428}{1000}.$$

Mais geralmente, mesmo que não termine em zeros, a expressão decimal $\alpha = a_0, a_1 a_2 \ldots a_n \ldots$ pode representar um número racional, desde que seja periódica. Comecemos com o caso mais simples, que é também o mais intrigante. Trata-se da expressão decimal, ou seja, do número real

$$\alpha = 0,999\ldots = \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \cdots$$

Afirmamos que $\alpha = 1$. Com efeito, os valores aproximados de α são $\alpha_1 = 0,9$, $\alpha_2 = 0,99$, $\alpha_3 = 0,999$, etc. Ora $1 - \alpha_1 = 0,1$, $1 - \alpha_2 = 0,01$, $1 - \alpha_3 = 0,001$ e, mais geralmente, $1 - \alpha_n = 10^{-n}$. Vemos portanto que, tomando n suficientemente grande, a diferença $1 - \alpha_n$ pode tornar-se tão pequena quanto se deseje. Noutras palavras, os números racionais $\alpha_n = 0,99\ldots9$ são valores cada vez mais aproximados de 1, ou seja, têm 1 como limite.

A igualdade $1 = 0,999\ldots$ costuma causar perplexidade aos menos experientes. A única maneira de dirimir o aparente paradoxo é esclarecer que o símbolo $0,999\ldots$ na realidade significa o número cujos valores aproximados são 0,9, 0,99, 0,999 etc. E, como vimos acima, esse número é 1.

Uma vez estabelecido que

$$0,999\ldots = \frac{9}{10} + \frac{9}{100} + \frac{9}{10^n} + \cdots = 1,$$

resulta imediatamente que

$$0,111\ldots = \frac{1}{10} + \frac{1}{100} + \cdots + \frac{1}{10^n} + \cdots = \frac{1}{9}.$$

Conseqüentemente, para todo dígito a, tem-se

$$0,aaa\ldots = \frac{a}{10} + \frac{a}{100} + \cdots + \frac{a}{10^n} + \cdots = \frac{a}{9}.$$

Por exemplo,

$$0,777\ldots = \frac{7}{9}.$$

Podemos ir mais além. Observando que

$$\frac{9}{10} + \frac{9}{100} = \frac{99}{100}, \frac{9}{1000} + \frac{9}{10000} = \frac{99}{10000}, \text{ etc.,}$$

obtemos:

$$\begin{aligned} 1 &= \left(\frac{9}{10} + \frac{9}{10^2}\right) + \left(\frac{9}{10^3} + \frac{9}{10^4}\right) + \cdots \\ &= \frac{99}{100} + \frac{99}{100^2} + \cdots \\ &= 99\left(\frac{1}{100} + \frac{1}{100^2} + \cdots\right), \end{aligned}$$

logo

$$\frac{1}{100} + \frac{1}{100^2} + \frac{1}{100^3} + \cdots = \frac{1}{99}.$$

Daí resulta, por exemplo, que

$$\begin{aligned} 0,373737\ldots &= \frac{37}{100} + \frac{37}{100^2} + \frac{37}{100^3} + \cdots \\ &= 37\left(\frac{1}{100} + \frac{1}{100^2} + \cdots\right) \\ &= \frac{37}{99}. \end{aligned}$$

Uma expressão decimal $\alpha = 0, a_1 a_3 \ldots$ chama-se uma *dízima periódica simples*, de período $a_1 a_2 \ldots a_n$, quando os primeiros p dígitos após a vírgula se repetem indefinidamente na mesma ordem. Assim, $0,777\ldots$ e $0,373737\ldots$ são dízimas periódicas simples, com períodos 7 e 37 respectivamente.

O raciocínio acima se aplica em geral e nos permite concluir que toda dízima periódica simples representa um número racional, que se chama sua *fração geratriz* (ou, simplesmente, sua *geratriz*). Mais precisamente, podemos dizer, como nos antigos compêndios de Aritmética:

A geratriz de uma dízima periódica simples é uma fração cujo numerador é o período e cujo denominador é o número formado por tantos noves quantos são os algarismos do período.

Por exemplo,

$$0,521521521\ldots = \frac{521}{999}.$$

Em particular, toda dízima periódica simples representa um número racional.

Existem ainda as dízimas periódicas ditas *compostas*. São aquelas que depois da vírgula têm uma parte que não se repete, seguida por uma parte periódica.

Para obter a geratriz de uma dízima periódica composta, procede-se como no exemplo a seguir:

$$\begin{aligned}
\alpha &= 0,35172172\ldots \\
100\alpha &= 35,172172 = 35\frac{172}{999} = \frac{35 \times 999 + 172}{999} = \\
&= \frac{35(1000-1)+172}{999} = \frac{35000+172-35}{999} = \frac{35172-35}{999}
\end{aligned}$$

Portanto

$$\alpha = \frac{35172-35}{99900}.$$

Chegamos assim à regra tradicional, que muitos de nós decoramos desde nossa infância:

A geratriz de uma dízima periódica composta é a fração cujo numerador é igual à parte não-periódica (35) seguida de um perí o-do (172) menos a parte não-periódica e cujo denominador é formado por tantos noves quantos são os algarismos do período, seguidos de tantos zeros quantos são os algarismos da parte não-periódica.

Em suma, expressões decimais periódicas (simples ou compostas) representam números racionais.

Reciprocamente, todo número racional é representado por uma expressão decimal finita (que acaba em zeros) ou periódica, como mostraremos a seguir.

A rigor, uma expressão decimal finita, como $0,35000\ldots$ é periódica, com período 0, mas é costume separar este caso, por ser muito particular.

Para obter a expressão decimal do número racional p/q, faz-se a "divisão continuada" de p por q, acrescentando-se zero ao dividendo p enquanto se tiver um resto não-nulo, como no exemplo abaixo

$$\frac{14}{27} = 0,518518\ldots \qquad \begin{array}{r|l} 140 & 27 \\ 050 & 0,518 \\ 230 & \\ 140 & \end{array}$$

Como nas divisões sucessivas só podem ocorrer os restos $0, 1, 2, \ldots$ $\ldots, q-1$, após no máximo q divisões um resto vai repetir-se e, a partir daí, os dígitos no quociente vão reaparecer na mesma ordem, logo tem-se uma expressão periódica.

Para um estudo mais detalhado sobre os casos em que o racional p/q gera uma dízima periódica simples, composta ou uma expressão decimal finita, bem como uma estimativa do número de algarimos do período, veja "Meu Professor de Matemática", págs. 158–171.

Observemos que a correspondência que associa a cada expressão decimal um número real é uma função sobrejetiva e "quase" injetiva.

A primeira das afirmações acima (sobrejetividade) significa

que, dado arbitrariamente um número real α, existe uma expressão decimal $a_0, a_1 a_2 \ldots a_n \ldots$ tal que $a_0 + a_1 \cdot 10^{-1} + a_2 \cdot 10^{-2} + \ldots + a_n \cdot 10^{-n} + \cdots = \alpha$. Como de costume, basta considerar o caso em que $\alpha > 0$. Então obtemos a expressão decimal de α tomando sucessivamente

a_0 = o maior número inteiro $\geqslant 0$ contido em (isto é, menor do que ou igual a) α;

a_1 = o maior dígito tal que $a_0 + \frac{a_1}{10} \leqslant \alpha$;

a_2 = o maior dígito tal que $a_0 + \frac{a_1}{10} + \frac{a_2}{100} \leqslant \alpha$;

e assim por diante.

Por exemplo, quando escrevemos que $\pi = 3,14159265\ldots$ estamos dizendo que

$$3 < \pi < 4, \qquad 3,1 < \pi < 3,2, \qquad 3,14 < \pi < 3,15 \quad \text{etc}$$

Quanto à "quase" injetividade da correspondência

$$\text{expressão decimal} \quad \mapsto \quad \text{número real},$$

o que estamos querendo dizer é que, se $0 \leqslant a_n \leqslant 8$ então as expressões decimais

$$a_0, a_1 \ldots a_n 999 \ldots \qquad \text{e} \qquad a_0, a_1 \ldots (a_n + 1)000 \ldots$$

definem o mesmo número real. Por exemplo,

$$3,275999\ldots = 3,276000\ldots$$
$$0,999\ldots = 1,000\ldots$$

A afirmação (um tanto imprecisa) de que uma correspondência é "quase" injetiva não tem sentido algum em geral. No presente caso, estamos querendo dizer que a situação acima descrita é a única em que há quebra de injetividade. Isto pode ser provado mas não haveria muita vantagem em fazê-lo aqui.

Para obter uma correspondência biunívoca entre as expressões decimais e os números reais, basta descartar as que terminam por uma seqüência de noves. Isto é o que faremos de agora em diante.

Operações com expressões decimais

Não é possível efetuar as quatro operações com as expressões decimais usando-as integralmente pois estas são organizadas da esquerda para a direita, enquanto as operações são normalmente desenvolvidas da direita para a esquerda. (Como começar uma adição, por exemplo?) Dados $\alpha = a_0, a_1a_2\ldots$ e $\beta = b_0, b_1b_2\ldots$, para calcular $\alpha + \beta$, $\alpha - \beta$, $\alpha \cdot \beta$ e α/β (se $\beta \neq 0$) toma-se $n \in \mathbb{N}$ e, considerando-se os valores aproximados $\alpha_n = a_0, a_1 \ldots a_n$, $\beta_n = b_0, b_1 \ldots b_n$, os números racionais $\alpha_n + \beta_n$, $\alpha_n - \beta_n$, $\alpha_n\beta_n$ e α_n/β_n são aproximações para os resultados que desejamos obter, tanto mais aproximados quanto maior for n.

Uma descoberta de George Cantor

Cantor foi a primeira pessoa a provar que existem diferentes números cardinais infinitos. Mais precisamente, os conjuntos $\mathbb{N}$ e $\mathbb{R}$ são ambos infinitos mas ele mostrou que não pode existir nenhuma função sobrejetiva $f: \mathbb{N} \to \mathbb{R}$. Em particular, não pode existir uma correspondência biunívoca entre $\mathbb{N}$ e $\mathbb{R}$. Como certamente existe uma função injetiva de $\mathbb{N}$ em $\mathbb{R}$ (a saber, aquela que a cada $n \in \mathbb{N}$ faz corresponder o próprio n, pensado como elemento de $\mathbb{R}$), diz-se então que a cardinalidade de $\mathbb{N}$ é estritamente menor do que a de $\mathbb{R}$.

A demonstração de Cantor consiste em mostrar que, dada qualquer função $f: \mathbb{N} \to \mathbb{R}$, é sempre possível achar um número real y que não pertence à imagem $f(\mathbb{N})$, isto é, tal que $f(n) \neq y$, seja qual for $n \in \mathbb{N}$.

Basta tomar um número real y cuja representação decimal tenha seu n-ésimo dígito diferente do n-ésimo dígito de $f(n)$, onde $n = 1, 2, 3, \ldots$. Isto garante que $y \neq f(n)$, seja qual for $n \in \mathbb{N}$, portanto $y \notin f(\mathbb{N})$.

Quando um conjunto é finito ou tem o mesmo número cardinal que $\mathbb{N}$, diz-se que ele é *enumerável*. O argumento de Cantor mostra que $\mathbb{R}$ não é enumerável. Ele também provou que o conjunto $\mathbb{Q}$ dos números racionais é enumerável. Não é difícil ver que a reunião de

dois conjuntos enumeráveis é ainda um conjunto enumerável. Se chamarmos de $\mathbb{Q}^c$ o conjunto dos números irracionais, teremos $\mathbb{R} = \mathbb{Q} \cup \mathbb{Q}^c$. Como $\mathbb{Q}$ é enumerável e $\mathbb{R}$ não é, resulta daí que o conjunto $\mathbb{Q}^c$ dos números irracionais é não-enumerável. Isto significa que existem muito mais números irracionais do que racionais! (Para mais detalhes, ver *Análise Real*, vol. 1, Capítulos 1 e 2.)

4. Desigualdades

A relação de desigualdade $x < y$ entre números reais é fundamental. Por isso é conveniente destacar algumas de suas propriedades, a fim de que saibamos o que estaremos fazendo quando operarmos com essa relação.

Em primeiro lugar, vale a pena lembrar que *todas* as propriedades das desigualdades derivam de duas afirmações simples e óbvias, que enunciaremos a seguir. Tais afirmações se referem aos números reais positivos. Para significar que o número real x é positivo, escreve-se $x > 0$. O conjunto dos números reais positivos será designado por $\mathbb{R}^+$. Assim

$$\mathbb{R}^+ = \{x \in \mathbb{R}; x > 0\}.$$

As propriedades básicas dos números positivos, das quais resulta tudo o que se pode provar sobre desigualdades, são as seguintes:

P1) Dado o número real x, há três possibilidades que se excluem mutuamente: ou x é positivo, ou $x = 0$ ou $-x$ é positivo.

P2) A soma e o produto de números positivos são ainda números positivos.

Com relação à propriedade P1), nunca é demais lembrar que $-x$ significa "x com o sinal trocado", ou seja, $-x$ é, por definição, o único número real tal que $-x + x = 0$.

Ainda com respeito a P1), quando $-x$ é positivo, diz-se que x é um número negativo e escreve-se $x < 0$.

A desigualdade entre números reais reduz-se ao conhecimento dos números positivos pois a afirmação $x < y$ significa que a

diferença $y - x$ é um número positivo. As propriedades essenciais da relação $x < y$ (que também se escreve $y > x$) são:

1) *Tricotomia*: dados $x, y \in \mathbb{R}$ vale uma, e somente uma, das alternativas seguintes: $x < y$, $x = y$ ou $y < x$.

2) *Transitividade*: se $x < y$ e $y < z$ então $x < z$.

3) *Monotonicidade da adição*: se $x < y$ então, para todo $z \in \mathbb{R}$ tem-se $x + z < y + z$.

4) *Monotonicidade da multiplicação*: se $x < y$ e z é positivo então $xz < yz$.

A tricotomia resulta imediatamente de P1). Com efeito, ou a diferença $y - x$ é positiva (em cujo caso $x < y$) ou é zero (e então $x = y$) ou é negativa (o que significa $y < x$).

Quanto à transitividade, ela se prova usando P2), assim: se $x < y$ e $y < z$ então $y - x$ e $z - y$ são positivos, logo a soma $z - x = (y - x) + (z - y)$ é positiva, ou seja, $x < z$.

A monotonicidade da adição, conforme está enunciada, segue-se imediatamente da definição de $x < y$. Com efeito, se $x < y$ então $y - x$ é positivo. Ora, $y - x = (y + z) - (x + z)$. Logo $x + z < y + z$. Há uma forma mais forte de enunciar a monotonicidade da adição, que é a seguinte:

3') Se $x < y$ e $x' < y'$ então $x + x' < y + y'$.

A propriedade 3') nos autoriza a somar *membro a membro* das desigualdades. Ela decorre de 2) e 3), assim:

Se $x < y$ e $x' < y'$ então, somando x' a ambos os membros da primeira igualdade e y a ambos os membros da segunda, em virtude de 3) obtemos $x + x' < y + x'$ e $y + x' < y + y'$. Por transitividade resulta então que $x + x' < y + y'$.

Finalmente, a monotonicidade da multiplicação resulta do fato de que o produto de dois números positivos é ainda um número positivo. Com efeito se $x < y$ e z é positivo então $y - x > 0$ e $z > 0$, logo $(y - x)z > 0$, ou seja $yz - xz > 0$, o que significa $xz < yz$.

Como no caso da adição, também é permitido multiplicar membro a membro duas desigualdades, desde que os números que nelas ocorrem sejam positivos. O enunciado preciso é:

4') Sejam x, y, x', y' números positivos. Se $x < y$ e $x' < y'$ então $xx' < yy'$.

Para provar isto, multiplicamos ambos os membros da desigualdade $x < y$ pelo número positivo x' e ambos os membros de $x' < y'$ pelo número positivo y, obtendo $xx' < yx'$ e $yx' < yy'$. Por transitividade, vem $xx' < yy'$.

As pessoas atentas a detalhes observarão que, para ser válida a propriedade 4'), basta que apenas três dos quatro números x, x', y e y' sejam positivos. (A demonstração acima requer apenas a positividade de x' e y mas, como $x' < y'$, daí resulta também que $y' > 0$.)

Outras propriedades que derivam de P1), P2) e suas conseqüências são:

5) Se $x \neq 0$ então $x^2 > 0$. (Todo quadrado, exceto 0, é positivo.)

Com efeito, se $x > 0$ então $x^2 > 0$ por P2). E se $-x > 0$ então, ainda por P2), $(-x)(-x) > 0$. Mas $(-x)(-x) = x^2$, logo $x^2 > 0$ em qualquer caso.

6) Se $0 < x < y$ então $0 < 1/y < 1/x$. (Quanto maior for um número positivo, menor será seu inverso.)

Em primeiro lugar, o inverso de um número positivo também é positivo porque $\frac{1}{x} = x \cdot \left(\frac{1}{x}\right)^2$ = produto de dois números positivos. (Veja 5).) Logo, multiplicando ambos os membros de $x < y$ pelo número positivo $1/xy$ vem $x/xy < y/xy$, isto é, $1/y < 1/x$.

7) Se $x < y$ e z é negativo então $xz > yz$. (Quando se multiplicam os dois membros de uma desigualdade por um número negativo, o sentido dessa desigualdade se inverte.)

Com efeito, o produto dos números positivos $y - x$ e $-z$ é positivo, isto é $(y - x)(-z) > 0$. Efetuando a multiplicação vem $xz - yz > 0$, portanto $xz > yz$.

A resolução de uma inequação com uma incógnita consiste na aplicação sucessiva das propriedades acima para simplificá-la até chegar a uma expressão final do tipo $x < c$ ou $x > c$.

Usa-se freqüentemente a notação $x \leqslant y$ para significar a nega-

ção de $y < x$. Portanto, $x \leqslant y$ significa que $x < y$ ou $x = y$. Por exemplo, são verdadeiras as afirmações $3 \leqslant 3$ e $5 \leqslant 7$.

Para encerrar estas considerações sobre desigualdade, lembraremos que a afirmação $x < y$, relativa aos números reais x e y, pode ser interpretada de três modos diferentes:

Geometricamente: $x < y$ significa que, num eixo orientado, o ponto de abcissa y está à direita do ponto de abcissa x.

Numericamente: Sejam

$$x = a_0, a_1 \ldots a_n \ldots \quad \text{e} \quad y = b_0, b_1 \ldots b_n \ldots$$

números reais positivos, dados por suas expressões decimais. Como se pode reconhecer que $x < y$? Certamente tem-se $x < y$ quando $a_0 < b_0$. (Lembre-se que estamos descartando as expressões decimais que terminam com uma seqüência de noves.) Ou então quando $a_0 = b_0$ e $a_1 < b_1$. Ou quando $a_0 = b_0$, $a_1 = b_1$ mas $a_2 < b_2$. E assim por diante. É como a ordem segundo a qual as palavras estão dispostas num dicionário. Tem-se $x < y$ se, e somente se, $a_0 < b_0$ ou então existe um inteiro $k > 0$ tal que $a_0 = b_0$, $a_1 = b_1, \ldots, a_{k-1} = b_{k-1}$ e $a_k < b_k$. Caso se tenha $x \leqslant 0 < y$, a relação $x < y$ é automática. E, finalmente, se x e y forem ambos negativos, tem-se $x < y$ se, e somente se, o número positivo $-y$ for menor do que o número positivo $-x$ segundo o critério acima.

Algebricamente: (Supondo conhecido o conjunto dos números positivos, gozando das propriedades P1) e P2) acima enunciados.) Tem-se $x < y$ se, e somente se, a diferença $d = y - x$ é um número positivo. Noutras palavras, vale $x < y$ se, e somente se, existe um número real positivo d tal que $y = x + d$.

Qual das três interpretações acima para o significado da desigualdade $x < y$ é a mais adequada? Todas são. As circunstâncias é que determinam qual é a mais conveniente.

5. Intervalos

Sejam a, b números reais, com $a \leqslant b$. Os nove subconjuntos de $\mathbb{R}$

abaixo definidos são chamados *intervalos*:

$$
\begin{aligned}
&[a,b] = \{x \in \mathbb{R}; a \leqslant x \leqslant b\}, && (-\infty, b] = \{x \in \mathbb{R}; x \leqslant b\},\\
&(a,b) = \{x \in \mathbb{R}; a < x < b\}, && (-\infty, b) = \{x \in \mathbb{R}; x < b\},\\
&[a,b) = \{x \in \mathbb{R}; a \leqslant x < b\}, && [a, +\infty) = \{x \in \mathbb{R}; a \leqslant x\},\\
&(a,b] = \{x \in \mathbb{R}; a < x \leqslant b\}, && (a, +\infty) = \{x \in \mathbb{R}; a < x\},\\
& && (-\infty, +\infty) = \mathbb{R}.
\end{aligned}
$$

Os quatro intervalos da esquerda são *limitados*, com extremos a, b : $[a,b]$ é um intervalo *fechado*, (a,b) é *aberto*, $[a,b)$ é *fechado à esquerda*, $(a,b]$ é *fechado à direita*. Os cinco intervalos da direita são *ilimitados*: $(-\infty, b]$ é a semi-reta esquerda, fechada, de origem b. Os demais têm denominações análogas. Quando $a = b$, o intervalo fechado $[a,b]$ reduz-se a um único elemento, chama-se um *intervalo degenerado* e os outros três intervalos da esquerda, neste caso, são vazios.

Alguns autores (principalmente os de livros escolares brasileiros) usam a notação $]a,b[$ em vez de (a,b) e, analogamente $[a,b[$, etc.

Deve-se ressaltar enfaticamente que $+\infty$ e $-\infty$ *não são números reais*. São apenas partes da notação de intervalos ilimitados.

Os intervalos são (com as notáveis exceções de $\mathbb{N}$, $\mathbb{Z}$ e $\mathbb{Q}$) os subconjuntos de $\mathbb{R}$ mais comumente encontrados.

Se tivéssemos de destacar um fato particularmente relevante a respeito de intervalos, provavelmente mencionaríamos o seguinte:

Todo intervalo não-degenerado contém números racionais e números irracionais.

Isto significa que os números racionais e os irracionais estão por toda parte em $\mathbb{R}$.

Para a demonstração da propriedade acima, veja *Análise Real*, vol. 1, pág. 19. O leitor pode, sem grande dificuldade, demonstrá-la usando a caracterização da desigualdade $a < b$ em termos das expressões decimais de a e b.

A propriedade acima destacada é essencial para provar o Teorema Fundamental da Proporcionalidade, do qual trataremos no capítulo seguinte.

Recomendação 2. A maioria de nossos livros escolares define número racional como "o número que pode ser expresso como quociente de dois inteiros", número irracional como "o número que não é racional" e $\mathbb{R}$ como o conjunto dos números racionais mais os irracionais. Como seus autores não dizem o que entendem por "número", resulta de suas definições que um número musical ou um número de uma revista são números irracionais. Não se deve adotar esse tipo de atitude. É verdade que a apresentação rigorosa da teoria dos números reais (conforme feita nos cursos de Análise) foge inteiramente ao nível e aos objetivos do ensino médio. Mas isto não deve ser motivo para escamoteações. Pelo contrário, quando se tem que falar sobre números reais para uma audiência matematicamente imatura, tem-se aí uma boa oportunidade para fazer a ligação entre a Matemática e o cotidiano, apresentando-os como resultados de medições, como tentamos explicar aqui.

6. Valor Absoluto

O *valor absoluto* (ou *módulo*) de um número real x, indicado pela notação $|x|$, é definido pondo-se

$$|x| = \begin{cases} x, & \text{se} \quad x \geqslant 0 \\ -x, & \text{se} \quad x < 0. \end{cases}$$

Outra maneira de se definir o valor absoluto consiste em pôr:

$$|x| = \max\{x, -x\},$$

isto é, o valor absoluto de x é maior dos números x e $-x$. (Quando $x = 0$ tem-se, é claro, $x = -x = |x| = 0$.)

Assim, por exemplo, $|x - 3| = x - 3$ se $x \geqslant 3$ e $|x - 3| = 3 - x$ quando $x \leqslant 3$.

Nas questões que envolvem o valor absoluto é-se, em princípio, obrigado a fazer as inevitáveis "considerações de casos", anali-

sando separadamente as situações conforme o sinal de cada expressão que ocorre no interior das barras verticais | |. Algumas vezes (infelizmente raras) isto pode ser evitado usando-se esta outra caracterização de valor absoluto:

$$|x| = \sqrt{x^2}.$$

Aqui estamos tirando partido da convenção que regula o uso do símbolo $\sqrt{\ }$: para todo $a \geqslant 0$, $\sqrt{a}$ é o número *não-negativo* cujo quadrado é a.

Outra importante interpretação do valor absoluto é a seguinte: se x e y são respectivamente as coordenadas dos pontos X e Y sobre o eixo E então

$$|x - y| = \text{distância do ponto X ao ponto Y}$$

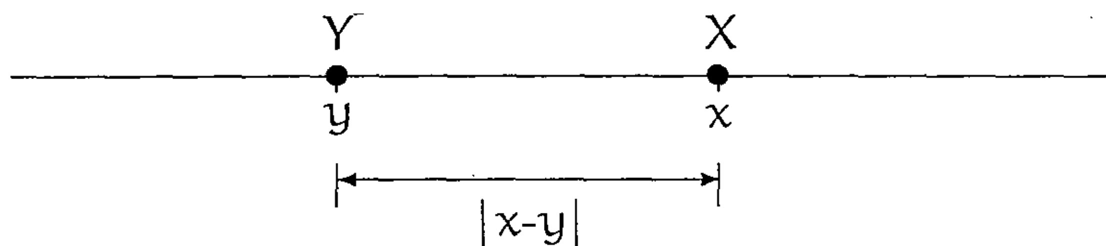


Figura 7

Para maiores detalhes sobre a distância entre dois pontos de um eixo, ver o livro "Coordenadas no Plano", página 5. As propriedades do valor absoluto são estudadas no livro "Análise Real", vol. 1, pag. 14.

A interpretação do valor absoluto $|x - y|$ como a distância, no eixo real, entre os pontos de coordenadas x e y, permite que se possa enxergar intuitivamente o significado e a resposta de algumas questões envolvendo módulos.

Por exemplo, a igualdade $|x - 2| = 3$ significa que o número x (ou o ponto que a ele corresponde no eixo) está a uma distância 3 do número 2. Logo, deve ser $x = 5$ (se x estiver à direita de 2) ou $x = -1$ (se estiver à esquerda).

Se tivermos uma desigualdade, como $|x - a| < \varepsilon$, com $\varepsilon > 0$, isto significa que a distância de x ao ponto a é menor do que ε, logo x

deve estar entre $a-\varepsilon$ e $a+\varepsilon$. Portanto o conjunto $\{x \in \mathbb{R}; |x-a| < \varepsilon\}$ é o intervalo aberto $(a-\varepsilon, a+\varepsilon)$.

Quando se lida com valores absolutos, não basta saber que $|x|$ é igual a x ou a $-x$. É necessário especificar quando é que se tem cada um desses casos. Esta observação deve ser aplicada especialmente na resolução de desigualdades.

7. Seqüências e Progressões

Uma *seqüência* é uma função cujo domínio é o conjunto $\mathbb{N}$ dos números naturais. Consideraremos apenas seqüências de números reais, isto é, funções de $\mathbb{N}$ em $\mathbb{R}$.

A notação usual para uma seqüência é $(x_1, x_2, \ldots, x_n, \ldots)$. Abreviadamente: $(x_n)_{n\in\mathbb{N}}$ ou (x_n), simplesmente. Isto significa que a seqüência dada é a função $1 \mapsto x_1, 2 \mapsto x_2, \ldots, n \mapsto x_n \ldots$, a qual faz corresponder a cada número natural n o número real x_n, chamado o n-*ésimo termo* da seqüência.

Exemplos particularmente interessantes de seqüências são as progressões.

Uma *progressão aritmética* (P.A.) é uma seqüência

$$x_1, x_2, \ldots, x_n, \ldots$$

onde cada termo, a partir do segundo, é a soma $x_{n+1} = x_n + r$ do termo anterior mais uma constante r, chamada a *razão* da progressão. Equivalentemente, a seqüência (x_n) chama-se uma progressão aritmética de razão r quando $x_{n+1} - x_n = r$ para todo $n \in \mathbb{N}$.

Na progressão aritmética (x_n) tem-se

$$x_2 = x_1 + r, \quad x_3 = x_2 + r = x_1 + 2r, \quad x_4 = x_1 + 3r, \ldots$$

e, em geral, $x_{n+1} = x_1 + nr$ para todo $n \in \mathbb{N}$.

A razão de uma progressão aritmética pode ser um número positivo, negativo ou igual a zero. No primeiro caso, a seqüência (x_n) é *crescente*, isto é, $m < n \Rightarrow x_m < x_n$. Quando a razão é negativa, a progressão aritmética é uma seqüência *decrescente*,

isto é, $m < n \Rightarrow x_n < x_m$. E, evidentemente, uma progressão aritmética de razão nula é constante: $x_1, x_1, x_1, \ldots$.

Uma *seqüência finita* (ou uma *lista*) é uma função cujo domínio tem a forma $I_n = \{1, 2, \ldots, n\}$. Ela é designada pela notação $(x_1, x_2, \ldots, x_n)$ e, neste caso, diz-se que se trata de uma seqüência com n termos. Em particular, uma seqüência (x_1, x_2) com dois termos é o que se chama um *par ordenado*.

Uma progressão aritmética finita (com n termos) é uma seqüência finita $(x_1, \ldots, x_n)$ tal que $x_2 - x_1 = x_3 - x_2 = \cdots = x_n - x_{n-1} = r$.

Uma progressão aritmética pode ser pensada como uma seqüência de pontos sobre uma reta, todos a igual distância dos seus vizinhos imediatos.

Uma *progressão geométrica* é uma seqüência

$$x_1, x_2, \ldots, x_n, \ldots$$

onde cada termo, a partir do segundo, é o produto $x_{n+1} = x_n \cdot r$ do anterior por uma constante r, chamada a *razão* da progressão. Tem-se portanto:

$$x_2 = x_1 \cdot r, \quad x_3 = x_2 \cdot r = x_1 \cdot r^2, \ldots, \text{ e, em geral, } x_{n+1} = x_1 \cdot r^n.$$

A igualdade $(1 - r)(1 + r + \cdots + r^{n-1}) = 1 - r^n$, de verificação imediata, mostra que a soma dos termos da progressão geométrica finita $1, r, r^2, \ldots, r^n$ é dada por

$$1 + r + r^2 + \cdots + r^n = \frac{1 - r^{n+1}}{1 - r} \quad \text{se} \quad r \neq 1.$$

Daí segue-se que, para uma progressão geométrica finita qualquer $x_1, x_2, \ldots, x_n$, de razão $r \neq 1$, tem-se

$$x_1 + x_2 + \cdots + x_n = x_1(1 + \cdots + r^{n-1}) = x_1 \frac{1 - r^n}{1 - r} = \frac{x_1 - x_n r}{1 - r}.$$

É tradicional e conveniente escrever P.A. e P.G. em vez de "progressão aritmética" e "progressão geométrica", respectivamente.

Exercícios

1. Dados os intervalos $A = [-1,3)$, $B = [1,4]$, $C = [2,3)$, $D = (1,2]$ e $E = (0,2]$ dizer se 0 pertence a $((A - B) - (C \cap D)) - E$.

2. Verifique se cada passo na solução das inequações abaixo está correto:

(a) $$\frac{5x+3}{2x+1} > 2 \quad \Rightarrow \quad 5x+3 > 4x+2 \quad \Rightarrow \quad x > -1$$

(b) $$\frac{2x^2+x}{x^2+1} < 2 \quad \Rightarrow \quad 2x^2+x < 2x^2+2 \quad \Rightarrow \quad x < 2$$

3. Seja $a, b, c, d > 0$ tais que $\frac{a}{b} < \frac{c}{d}$. Mostre que

$$\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}.$$

Interprete este resultado no caso em que a, b, c e d são inteiros positivos (isto é, o que significa somar numeradores e denominadores de duas frações?)

4. Qual é a aproximação de raiz cúbica de 3 por falta com uma casa decimal?

5. Ao terminar um problema envolvendo radicais, os alunos normalmente são instados a racionalizar o denominador do resultado obtido. Por que isso?

6. Considere todos os intervalos da forma $[0, \frac{1}{n}]$. Existe um número comum a todos estes intervalos? E se forem tomados os intervalos abertos?

7. Considere um número racional m/n, onde m e n são primos entre si. Sob que condições este número admite uma representação decimal finita? Quando a representação é uma dízima periódica simples?

8. O número $0,12345678910111213141 5\ldots$ é racional ou irracional?

9. Utilize a interpretação geométrica de módulo para resolver as equações e inequações abaixo:

a) $|x-1|=4$

b) $|x+1|<2$

c) $|x-1|<|x-5|$

d) $|x-2|+|x+4|=8$

e) $|x-2|+|x+4|=1$

10. Sejam a e b números reais não negativos. Mostre que

$$\left(\frac{a+b}{2}\right)^2 \leq \frac{a^2+b^2}{2}.$$

Interprete geometricamente esta desigualdade.

11. Sabendo que os números reais x, y satisfazem as desigualdades $1,4587 < x < 1,4588$ e $0,1134 < y < 0,1135$, têm-se os valores exatos de x e y até milésimos. Que grau de precisão, a partir daí, podemos ter para o valor de xy? Determine esse valor aproximado. Como procederíamos para obter um valor aproximado de x/y? Qual o grau de precisão encontrado no caso do quociente?

Capítulo 5

Funções Afins

O assunto principal deste capítulo e dos seguintes são as funções reais de uma variável real, isto é, funções $f: X \to \mathbb{R}$ que têm como domínio um subconjunto $X \subset \mathbb{R}$ e cujos valores $f(x)$, para todo $x \in X$, são números reais. Em cada um desses capítulos, abordaremos um tipo particular de função, começando com o caso mais simples e aumentando pouco a pouco a complexidade.

Iniciaremos com a função afim, cujo estudo será precedido de uma breve revisão sobre o produto cartesiano e o gráfico de uma função.

0. Produto Cartesiano

Um *par ordenado* $p = (x, y)$ é formado por um objeto x, chamado a *primeira coordenada* de p e um objeto y, chamado a *segunda coordenada* de p. Dois pares ordenados $p = (x, y)$ e $q = (u, v)$ serão chamados iguais quando $x = u$ e $y = v$, isto é, quando tiverem a mesma primeira coordenada e a mesma segunda coordenada.

É permitido considerar o par ordenado (x, x), no qual a primeira coordenada coincide com a segunda.

O par ordenado $p = (x, y)$ não é a mesma coisa que o conjunto $\{x, y\}$ porque $\{x, y\} = \{y, x\}$ sempre, mas $(x, y) = (y, x)$ somente quando $x = y$.

O *produto cartesiano* $X \times Y$ de dois conjuntos X e Y é o conjunto $X \times Y$ formado por todos os pares ordenados (x, y) cuja primeira coordenada x pertence a X e cuja segunda coordenada y pertence

a Y. Simbolicamente:

$$X \times Y = \{(x, y); x \in X, y \in Y\}.$$

Se $X = \{x_1, \ldots, x_m\}$ e $Y = \{y_1, \ldots, y_p\}$ são conjuntos finitos com m e p elementos respectivamente, então o produto cartesiano $X \times Y$ é finito e possui mp elementos. Noutras palavras, $n(X \times Y) = n(X) \cdot n(Y)$. A melhor maneira de enxergar isto é pensar no produto cartesiano $X \times Y$ como um quadro retangular

$$\begin{matrix} (x_1, y_1)(x_1, y_2) \ldots (x_1, y_p) \\ (x_2, y_1)(x_2, y_2) \ldots (x_2, y_p) \\ \vdots \\ (x_m, y_1)(x_m, y_2) \ldots (x_m, y_p) \end{matrix}$$

com p colunas, cada uma das quais possui m elementos.

Exemplo 1. Sejam AB e CD segmentos de reta. O produto cartesiano $AB \times CD$ pode ser interpretado como um retângulo, na forma indicada pela figura. Tomamos AB e CD perpendiculares e cada elemento $(x, y) \in AB \times CD$ é representado pelo ponto P, interseção das perpendiculares a AB e CD tiradas pelos pontos x e y respectivamente.

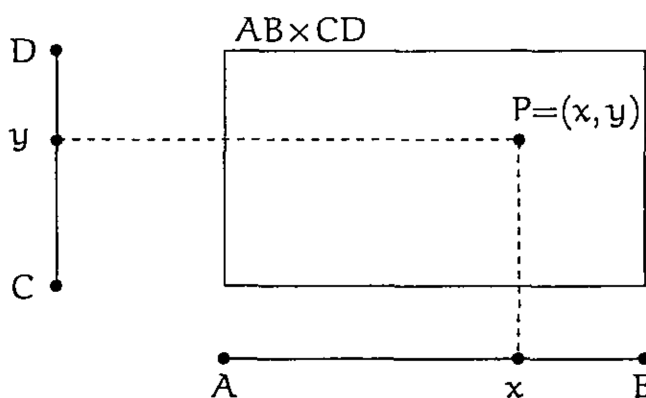


Figura 8

Exemplo 2. Na mesma veia do exemplo anterior, o produto cartesiano $\gamma \times AB$ de uma circunferência γ por um segmento de reta AB é representado por um cilindro.

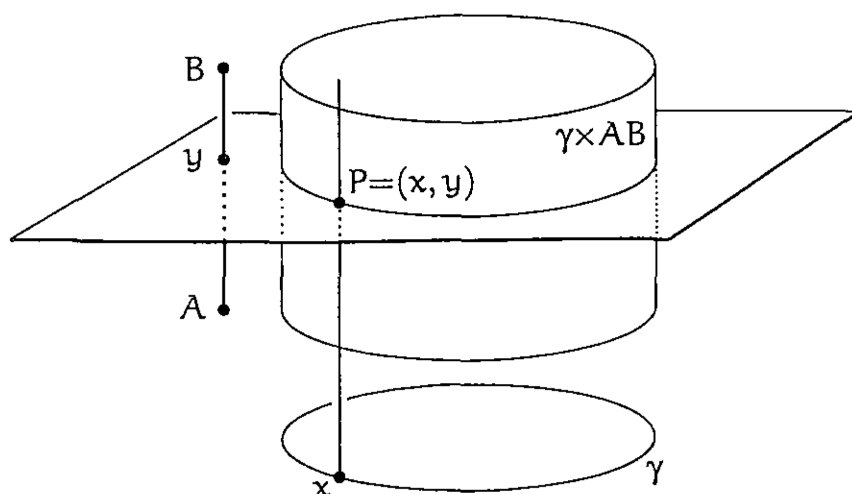


Figura 9

Para isto, tomamos o segmento AB perpendicular ao plano de γ. Cada elemento (x, y) do produto cartesiano $\gamma \times AB$ é representado pelo ponto P, interseção da reta perpendicular ao plano de γ tirada pelo ponto x com o plano perpendicular ao segmento AB tirado pelo ponto y.

O *gráfico* de uma função $f: X \rightarrow Y$ é o subconjunto $G(f)$ do produto cartesiano $X \times Y$ formado por todos os pares ordenados (x, y), onde x é um ponto qualquer de X e $y = f(x)$. Assim,

$$G(f) = \{(x, y) \in X \times Y; y = f(x)\} = \{(x, f(x)); x \in X\}.$$

A fim de que um subconjunto $G \subset X \times Y$ seja o gráfico de alguma função $f: X \rightarrow Y$ é necessário e suficiente que G cumpra as seguintes condições:

G1. Para todo $x \in X$ existe um par ordenado $(x, y) \in G$ cuja primeira coordenada é x.

G2. Se $p = (x, y)$ e $p' = (x, y')$ são pares pertencentes a G com a mesma primeira coordenada x então $y = y'$ (isto é, $p = p'$).

É claro que estas condições podem ser resumidas numa só, dizendo-se que para cada $x \in X$ existe um, e somente um, $y \in Y$ tal que $(x, y) \in G$.

O produto cartesiano $X \times Y$ acha-se intimamente ligado à idéia de relação ou, mais precisamente, relação binária. Uma *relação* (binária) R entre elementos do conjunto X e elementos do conjunto

Y é uma condição ou um conjunto de condições que permitem determinar, dados $x \in X$ e $y \in Y$, se x está ou não relacionado com y segundo R. No caso afirmativo, escreve-se $x R y$.

Um exemplo à mão é a relação "menor do que" entre números reais. A condição que nos permite escrever $x < y$, com $x \in \mathbb{R}$ e $y \in \mathbb{R}$ é $y - x > 0$. Trata-se aqui de uma relação entre $\mathbb{R}$ e $\mathbb{R}$. Para outro exemplo, consideramos o conjunto D de todas as retas e o conjunto P de todos os planos do espaço. O paralelismo entre uma reta r e um plano Π é uma relação entre elementos de D e elementos de P que se escreve $r \parallel \Pi$ e significa que a reta r e o plano Π não têm elementos em comum.

Um exemplo particularmente importante de relação é a relação funcional. Ela ocorre quando se tem uma função $f: X \to Y$. Diz-se então que o elemento $x \in X$ está relacionado com o elemento $y \in Y$ quando $y = f(x)$. Neste caso, não se costuma escrever $x f y$ como se faria numa outra relação qualquer. Põe-se apenas $y = f(x)$.

O *gráfico* de uma relação R entre os conjuntos X e Y é o subconjunto $G(R)$ do produto cartesiano $X \times Y$ formado pelos pares (x, y) tais que $x R y$. Assim, $G(R) = \{(x, y) \in X \times Y; x R y\}$.

Esta noção inclui o caso particular do gráfico de uma função.

Recomendação 1. Praticamente todos os textos escolares em uso no nosso país definem uma função $f: X \to Y$ como um subconjunto do produto cartesiano $X \times Y$ com as propriedades G1 e G2 acima enunciadas. Essa definição apresenta o inconveniente de ser formal, estática e não transmitir a idéia intuitiva de função como correspondência, transformação, dependência (uma grandeza função de outra) ou resultado de um movimento. Quem pensaria numa rotação como um conjunto de pares ordenados? Os matemáticos e (principalmente) os usuários da Matemática olham para uma função como uma correspondência, não como um conjunto de pares ordenados. Poder-se-ia talvez abrir uma exceção para os lógicos, quando querem mostrar que todas as noções matemáticas se reduzem, em última análise, à idéia pura de conjunto.

Mas certamente este não é o caso aqui. Se definimos uma função $f: X \to Y$ como um subconjunto particular do produto cartesiano $X \times Y$, qual seria a definição matemática do gráfico de uma função?

Em suma, a terminologia que consideramos adequada é a seguinte: um subconjunto qualquer de $X \times Y$ é o *gráfico* de uma relação de X para Y. Se esse conjunto cumpre as condições G1 e G2 acima estipuladas, ele é o *gráfico* de uma função.

1. O Plano Numérico $\mathbb{R}^2$

$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ é o exemplo mais importante de produto cartesiano pois, afinal de contas, trata-se do caso particular que deu origem à idéia geral.

Os elementos (x, y) de $\mathbb{R}^2$ são, naturalmente, os pares ordenados de números reais. Eles surgem como as coordenadas cartesianas de um ponto P do plano Π (x = abcissa, y = ordenada) quando se fixa nesse plano um par de eixos ortogonais OX e OY, que se intersectam no ponto O, chamado a *origem* do sistema de coordenadas.

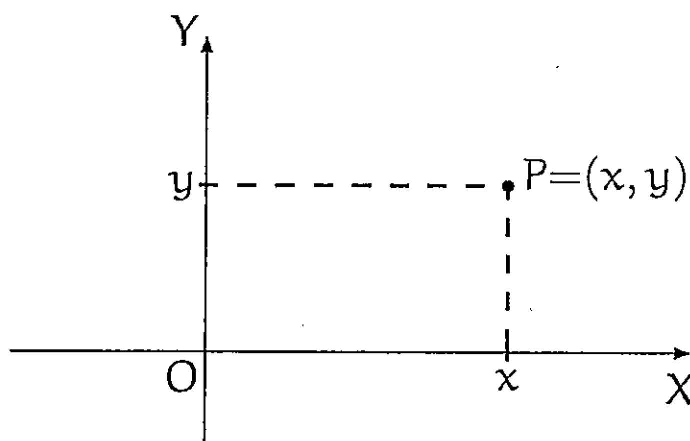


Figura 10

Dado o ponto $P \in \Pi$, a *abcissa* de P é o número x, coordenada do pé da perpendicular baixada de P sobre o eixo OX, enquanto a *ordenada* de P é a coordenada y do pé da perpendicular baixada de P sobre o eixo OY. Diz-se então que (x, y) é o par de *coordenadas* do ponto P relativamente ao sistema de eixos OXY. Os eixos OX

e OY dividem o plano em quatro regiões, chamadas *quadrantes*, caracterizadas pelos sinais das coordenadas de seus pontos. No primeiro quadrante, tem-se $x \geqslant 0$ e $y \geqslant 0$; no segundo, $x \leqslant 0$ e $y \geqslant 0$; no terceiro, $x \leqslant 0$ e $y \leqslant 0$; no quarto, $x \geqslant 0$ e $y \leqslant 0$.

A função $f \colon \Pi \to \mathbb{R}^2$, que associa a cada ponto P do plano Π seu par de coordenadas $f(P) = (x, y)$ relativamente ao sistema de eixos OXY, é uma correspondência biunívoca. Ela permite traduzir conceitos e propriedades geométricas para uma linguagem algébrica e, reciprocamente, interpretar geometricamente relações entre números reais.

Podemos então dizer que $\mathbb{R}^2$ é o modelo aritmético do plano Π enquanto Π é o modelo geométrico de $\mathbb{R}^2$.

Do nosso presente ponto de vista, olharemos para $\mathbb{R}^2$ como um plano, (o plano numérico), chamaremos seus elementos $P = (x, y)$ de *pontos* e procuraremos, com ajuda dessa linguagem geométrica e dos resultados da Geometria, alcançar um melhor entendimento das propriedades das funções reais que vamos estudar. Veremos pouco a pouco as vantagens desse caminho de mão dupla que liga a Aritmética e a Álgebra de um lado à Geometria do outro.

A pergunta mais básica, uma das primeiras que se impõe responder, é a seguinte: se $P = (x, y)$ e $Q = (u, v)$, como se pode exprimir a distância do ponto P ao ponto Q em termos dessas coordenadas?

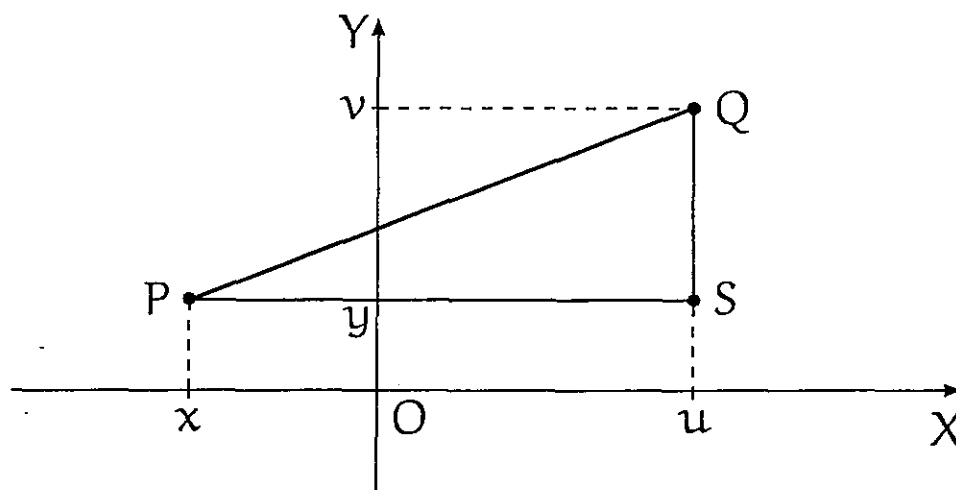


Figura 11

A resposta é fornecida imediatamente pelo Teorema de Pitágoras. Introduzimos o ponto auxiliar $S = (u, y)$.

Como P e S têm a mesma ordenada, o segmento PS é horizontal (paralelo ao eixo OX). Analogamente, QS é vertical (paralelo a OY). Portanto o segmento PQ é a hipotenusa do triângulo retângulo PQS, cujos catetos medem $|x - u|$ e $|y - v|$ respectivamente. (Vide seção 6 do Capítulo 4.) O Teorema de Pitágoras nos dá então:

$$d(P, Q)^2 = (x - u)^2 + (y - v)^2,$$

ou seja:

$$d(P, Q) = \sqrt{(x - u)^2 + (y - v)^2}.$$

Em particular, a distância do ponto $P = (x, y)$ à origem $O = (0, 0)$ é igual a

$$\sqrt{x^2 + y^2}.$$

Exemplo 3. Se o centro de uma circunferência C é o ponto $A = (a, b)$ e o raio é o número real $r > 0$ então, por definição, um ponto $P = (x, y)$ pertence a C se, e somente se, $d(A, P) = r$. Pela fórmula da distância entre dois pontos, vemos que

$$C = \{(x, y); (x - a)^2 + (y - b)^2 = r^2\}.$$

Diz-se então que

$$(x - a)^2 + (y - b)^2 = r^2$$

é a equação da circunferência de centro no ponto $A = (a, b)$ e raio r.

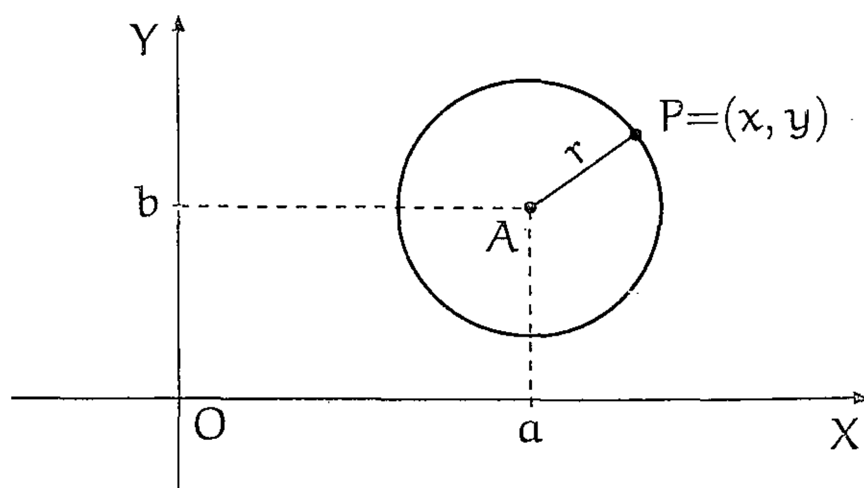


Figura 12

Por sua vez, o *disco* D de centro A e raio r é formado pelos pontos $P = (x, y)$ cuja distância ao ponto A é $\leqslant r$. Portanto

$$D = \{(x, y); (x - a)^2 + (y - b)^2 \leqslant r^2\}.$$

Recomendação 2. A palavra *círculo* é ambígua. Às vezes significa a circunferência, às vezes quer dizer o disco que tem essa circunferência como fronteira. Não é errado usá-la com qualquer desses dois significados. (Euclides já o fazia. Além disso, os termos polígono, elipse, triângulo, quadrado, etc. também têm duplo sentido.) Mas é necessário explicar o que se está querendo dizer, para evitar mal-entendidos.

O gráfico de uma função real de variável real $f: X \to \mathbb{R}$ é um subconjunto do plano numérico $\mathbb{R}^2$, logo pode ser visualizado (pelo menos nos casos mais simples) como uma linha, formada pelos pontos de coordenadas $(x, f(x))$, quando x varia no conjunto X.

Exemplo 4. A fórmula da distância entre dois pontos serve para reconhecer que o gráfico G da função $f: [-1, 1] \to \mathbb{R}$, dada por

$$f(x) = \sqrt{1 - x^2},$$

é a semi-circunferência C_+, de centro na origem $= (0, 0)$ e raio 1, situada no semi-plano $y \geqslant 0$

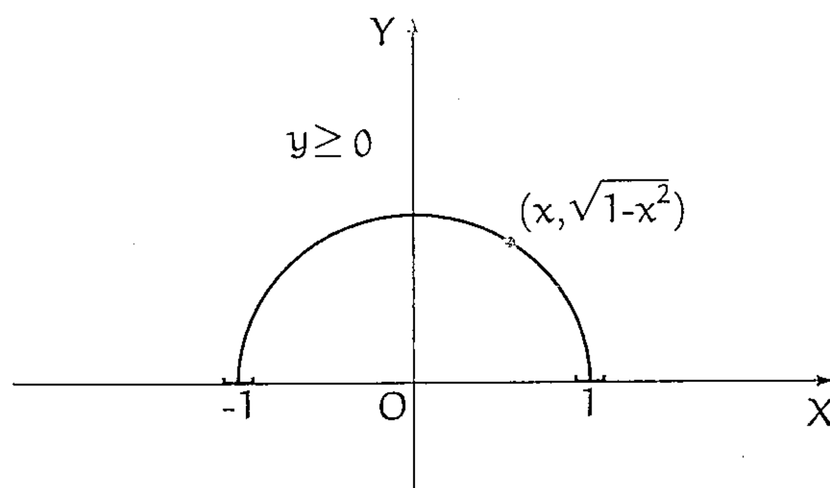


Figura 13

Com efeito,

$$\begin{aligned}(x,y) \in G &\Leftrightarrow -1 \leqslant x \leqslant 1 \quad \text{e} \quad y = \sqrt{1-x^2}\\ &\Leftrightarrow -1 \leqslant x \leqslant 1, y \geqslant 0 \quad \text{e} \quad y^2 = 1 - x^2\\ &\Leftrightarrow y \geqslant 0 \quad \text{e} \quad x^2 + y^2 = 1\\ &\Leftrightarrow (x,y) \in C_+ .\end{aligned}$$

No caso de funções reais de uma variável real, as condições G1 e G2 adquirem uma forma mais geométrica e são resumidas assim:

Seja $X \subset \mathbb{R}$ um conjunto que consideraremos situado sobre o eixo horizontal. Um subconjunto $G \subset \mathbb{R}^2$ é o gráfico de uma função $f\colon X \to \mathbb{R}$ se, e somente se, toda reta paralela ao eixo vertical, traçada a partir de um ponto de X, intersecta G num único ponto.

Exemplo 5. Dado o número real $c \neq 0$, consideremos o conjunto G, formado pelos pontos (x, y) de $\mathbb{R}^2$ tais que $xy = c$. Simbolicamente,

$$G = \{(x,y) \in \mathbb{R}^2; xy = c\}.$$

O conjunto G é o que se chama uma *hipérbole equilátera*. A figura abaixo mostra a forma de G nos casos $c > 0$ e $c < 0$. Para todo $x \neq 0$, a reta vertical que passa pelo ponto de abcissa x corta o conjunto G no único ponto $(x, c/x)$. Logo, G é o gráfico da função $f\colon \mathbb{R} - \{0\} \to \mathbb{R}$, dada por $f(x) = c/x$.

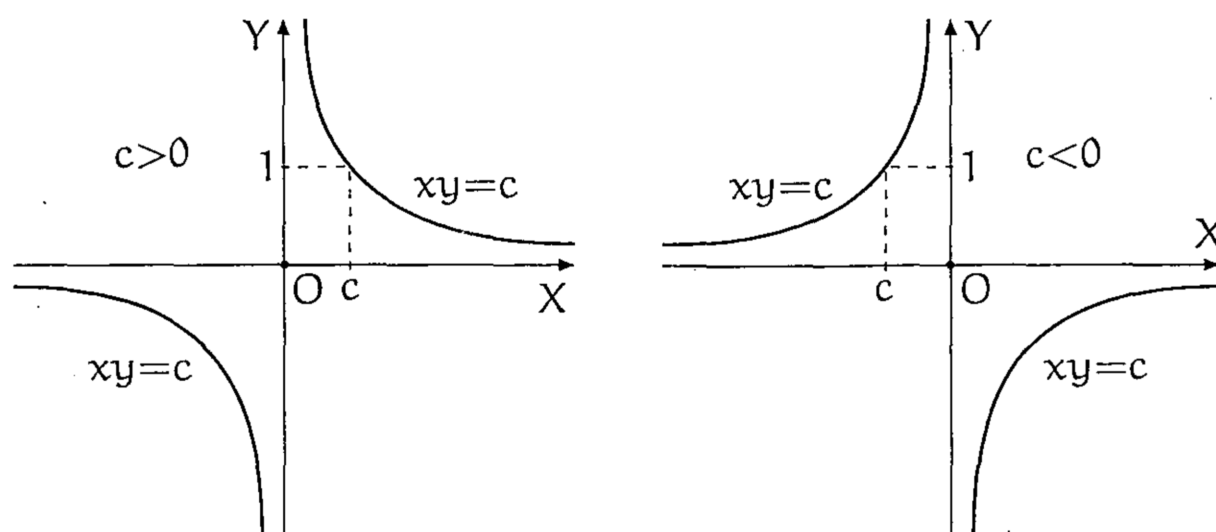


Figura 14

2. A Função Afim

Uma função $f: \mathbb{R} \to \mathbb{R}$ chama-se *afim* quando existem constantes $a, b \in \mathbb{R}$ tais que $f(x) = ax + b$ para todo $x \in \mathbb{R}$.

Exemplo 6. A *função identidade* $f: \mathbb{R} \to \mathbb{R}$, definida por $f(x) = x$ para todo $x \in \mathbb{R}$, é afim. Também são afins as *translações* $f: \mathbb{R} \to \mathbb{R}$, $f(x) = x + b$. São ainda casos particulares de funções afins as funções *lineares*, $f(x) = ax$ e as funções *constantes* $f(x) = b$.

É possível, mediante critérios como os que apresentaremos logo a seguir, saber que uma certa função $f: \mathbb{R} \to \mathbb{R}$ é afim sem que os coeficientes a e b sejam fornecidos explicitamente. Neste caso, obtém-se b como o valor que a função dada assume quando $x = 0$. O número $b = f(0)$ às vezes se chama o *valor inicial* da função f. Quanto ao coeficiente a, ele pode ser determinado a partir do conhecimento dos valores $f(x_1)$ e $f(x_2)$ que a função f assume em dois pontos distintos (porém arbitrários) x_1 e x_2. Com efeito, conhecidos

$$f(x_1) = ax_1 + b$$

e

$$f(x_2) = ax_2 + b,$$

obtemos

$$f(x_2) - f(x_1) = a(x_2 - x_1),$$

portanto

$$a = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

Dados $x, x+h \in \mathbb{R}$, com $h \neq 0$, o número $a = [f(x+h) - f(x)]/h$ chama-se a *taxa de crescimento* (ou taxa de variação) da função f no intervalo de extremos x, $x+h$.

Lembremos que uma função $f: X \to \mathbb{R}$, com $X \subset \mathbb{R}$, chama-se:
crescente quando $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$;
decrescente quando $x_1 < x_2 \Rightarrow f(x_1) > f(x_2)$;
monótona não-decrescente quando $x_1 < x_2 \Rightarrow f(x_1) \leqslant f(x_2)$;
monótona não-crescente quando $x_1 < x_2 \Rightarrow f(x_1) \geqslant f(x_2)$.

Em qualquer dos quatro casos, f diz-se *monótona*. Nos dois primeiros (f crescente ou f decrescente) diz-se que f é *estritamente monótona*. Nestes dois casos, f é uma função injetiva.

Recomendação 3. Não fica bem (embora algumas vezes se faça) chamar apenas de não-decrescentes e não-crescentes as funções dos dois últimos tipos, pois negar (por exemplo) que uma função seja decrescente não implica necessariamente que ela seja monótona.

Evidentemente, os quatro casos acima não são mutuamente excludentes. Pelo contrário, os dois primeiros são casos particulares dos dois últimos. Além disso, naturalmente, há funções que não se enquadram em nenhuma dessas quatro categorias.

Uma função afim é crescente quando sua taxa de crescimento (o coeficiente a) é positiva, decrescente quando a é negativo e constante quando $a = 0$.

Exemplo 7. O preço a pagar por uma corrida de taxi é dado por uma função afim $f: x \mapsto ax + b$, onde x é a distância percorrida (usualmente medida em quilômetros), o valor inicial b é a chamada *bandeirada* e o coeficiente a é o preço de cada quilômetro rodado.

O gráfico G *de uma função afim* $f: x \mapsto ax+b$ *é uma linha reta.*

Para ver isto basta mostrar que três pontos quaisquer

$$P_1 = (x_1, ax_1 + b),$$
$$P_2 = (x_2, ax_2 + b) \quad \text{e}$$
$$P_3 = (x_3, ax_3 + b)$$

desse gráfico são colineares. Para que isto ocorra, é necessário e suficiente que o maior dos três números $d(P_1, P_2)$, $d(P_2, P_3)$ e $d(P_1, P_3)$ seja igual à soma dos outros dois. Ora, podemos sempre supor que as abcissas x_1, x_2 e x_3 foram numeradas de modo que $x_1 < x_2 < x_3$. A fórmula da distância entre dois pontos nos dá:

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + a^2(x_2 - x_1)^2}$$
$$= (x_2 - x_1)\sqrt{1 + a^2},$$
$$d(P_2, P_3) = (x_3 - x_2)\sqrt{1 + a^2}$$

e

$$d(P_1, P_3) = (x_3 - x_1)\sqrt{1 + a^2}.$$

Daí se segue imediatamente que

$$d(P_1, P_3) = d(P_1, P_2) + d(P_2, P_3).$$

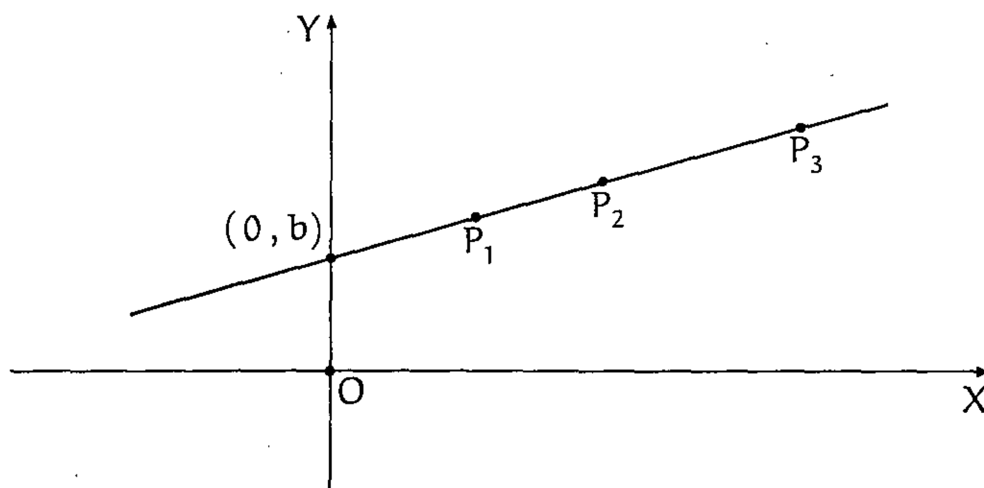


Figura 15

Do ponto de vista geométrico, b é a ordenada do ponto onde a reta, que é o gráfico da função $f: x \mapsto ax + b$, intersecta o eixo OY.

O número a chama-se *a inclinação*, ou *coeficiente angular*, dessa reta (em relação ao eixo horizontal OX). Quanto maior o valor de a mais a reta se afasta da posição horizontal. Quando $a > 0$, o gráfico de f é uma reta ascendente (quando se caminha para a direita) e quando $a < 0$, a reta é descendente.

De acordo com a letra estrita da definição, a fim de conhecer uma função $f: X \to Y$, deve-se ter uma regra que permita (pelo menos teoricamente) determinar o valor $f(x)$ *para todo* $x \in X$.

No caso particular de uma função afim $f: \mathbb{R} \to \mathbb{R}$, como seu gráfico é uma linha reta e como uma reta fica inteiramente determinada quando se conhecem dois de seus pontos, resulta que basta conhecer os valores $f(x_1)$ e $f(x_2)$, que a função afim $f: \mathbb{R} \to \mathbb{R}$ assume em dois números $x_1 \neq x_2$ (escolhidos arbitrariamente), para que f fique inteiramente determinada.

Na prática, sabendo que $f: \mathbb{R} \to \mathbb{R}$ é afim e que $f(x_1) = y_1$, $f(x_2) = y_2$ com $x_1 \neq x_2$, queremos determinar os coeficientes a e b de modo que se tenha $f(x) = ax + b$ para todo $x \in \mathbb{R}$. Isto corresponde a resolver o sistema

$$ax_1 + b = y_1$$
$$ax_2 + b = y_2,$$

no qual as incógnitas são a e b (!). A solução é imediata:

$$a = \frac{y_2 - y_1}{x_2 - x_1}, \qquad b = \frac{x_2 y_1 - x_1 y_2}{x_2 - x_1}.$$

[Em geral, sempre que precisamos fazer a hipótese $x_1 \neq x_2$ para resolver um problema, a diferença $x_2 - x_1$ costuma aparecer em algum denominador na solução.]

O argumento acima provou que

Dados arbitrariamente (x_1, y_1), $(x_2, y_2) \in \mathbb{R}^2$, *com* $x_1 \neq x_2$, *existe uma, e somente uma, função afim* $f: \mathbb{R} \to \mathbb{R}$ *tal que* $f(x_1) = y_1$ *e* $f(x_2) = y_2$.

Evidentemente, o gráfico de uma função afim é uma reta não-vertical, isto é, não paralela ao eixo OY. Reciprocamente:

Toda reta não-vertical r é o gráfico de uma função afim.

Para provar esta afirmação, tomemos dois pontos distintos $P_1 = (x_1, y_1)$ e $P_2 = (x_2, y_2)$ na reta r. Como r não é vertical, temos necessariamente $x_1 \neq x_2$, logo existe uma função afim $f: \mathbb{R} \to \mathbb{R}$ tal que $f(x_1) = y_1$ e $f(x_2) = y_2$. O gráfico de f é uma reta que passa pelos pontos P_1 e P_2 logo essa reta coincide com r.

Se $f(x) = ax + b$, diz-se que $y = ax + b$ é a *equação* da reta r.

Se a reta r é o gráfico da função afim f, dada por $f(x) = ax + b$, o coeficiente

$$a = \frac{y_2 - y_1}{x_2 - x_1},$$

onde (x_1, y_1) e (x_2, y_2) são dois pontos distintos quaisquer de r, tem claramente o significado de taxa de crescimento de f. A esse número é dado também o nome de inclinação ou coeficiente angular da reta r, pois ele é a tangente trigonométrica do ângulo do eixo OX com a reta r.

Estas interpretações nos levam a concluir imediatamente que a equação da reta que passa pelos pontos (x_1, y_1) e (x_2, y_2), não situados na mesma vertical é

$$y = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

ou

$$y = y_2 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_2).$$

(Os segundos membros destas equações são iguais!) A primeira equação nos diz que, se começarmos no ponto (x_1, y_1) e caminharmos sobre a reta, fazendo x variar, a ordenada y começa com o valor y_1 e sofre um incremento igual ao incremento $x - x_1$ dado a x, vezes a taxa de variação

$$\frac{y_2 - y_1}{x_2 - x_1}.$$

A segunda equação diz a mesma coisa, só que partindo do ponto (x_2, y_2).

De modo análogo, vemos que a equação da reta que passa pelo ponto (x_0, y_0) e tem inclinação a é

$$y = y_0 + a(x - x_0).$$

Comentários sobre terminologia

1. Se a função afim f é dada por $f(x) = ax + b$, não é adequado chamar o número a de *coeficiente angular* da função f. O nome mais apropriado, que usamos, é *taxa de variação* (ou taxa de crescimento). Em primeiro lugar não há, na maioria dos casos, ângulo algum no problema estudado. Em segundo lugar, mesmo considerando o gráfico de f, o ângulo que ele faz com o eixo horizontal depende das unidades escolhidas para medir as grandezas x e $f(x)$. Em resumo: tem-se taxa de variação de uma função e coeficiente angular de uma reta.

2. A maioria dos nossos textes escolares refere-se à função afim como "função do primeiro grau". Essa nomenclatura sugere a pergunta: o que é o grau de uma função? Função não tem grau. O que possui grau é um polinômio. (Quando $a \neq 0$, a expressão $f(x) = ax + b$ é um polinômio do primeiro grau.) O mesmo defeito de nomenclatura ocorre também com as funções quadráticas, que estudaremos no capítulo seguinte. Elas muitas vezes são chamadas, incorretamente, "funções do segundo grau".

3. A Função Linear

A função linear, dada pela fórmula $f(x) = ax$, é o modelo matemático para os problemas de proporcionalidade. A proporcionalidade é, provavelmente, a noção matemática mais difundida na cultura de todos os povos e seu uso universal data de milênios.

Sem ir tão longe, vejamos como este assunto era tratado em nosso país pelas gerações que nos antecederam. Para isto, vamos consultar um compêndio antigo e muito bem conceituado,

sem dúvida o texto matemático de mais longa utilização no Brasil. Trata-se da *Aritmética Progressiva*, de Antonio Trajano, cuja primeira edição ocorreu em 1883 e ainda se achava em circulação na década de 60, com mais de oitenta edições publicadas. Trajano dá a seguinte definição:

Diz-se que duas grandezas são proporcionais quando elas se correspondem de tal modo que, multiplicando-se uma quantidade de uma delas por um número, a quantidade correspondente da outra fica multiplicada ou dividida pelo mesmo número. No primeiro caso, a proporcionalidade se chama direta e, no segundo, inversa; as grandezas se dizem diretamente proporcionais ou inversamente proporcionais.

Substituindo as grandezas de Trajano por suas medidas, que são números reais, podemos traduzir o que está dito acima para nossa linguagem atual, da seguinte forma:

Uma proporcionalidade é uma função $f\colon \mathbb{R} \to \mathbb{R}$ *tal que, para quaisquer números reais* c, x *tem-se* $f(cx) = c \cdot f(x)$ *(proporcionalidade direta) ou* $f(cx) = f(x)/c$, *se* $c \neq 0$ *(proporcionalidade inversa).*

Nesta nova versão, as grandezas da definição antiga são os números reais x, y e a corresponência a que Trajano se refere é uma função $f\colon \mathbb{R} \to \mathbb{R}$ tal que $y = f(x)$.

É claro que se $f(cx) = c \cdot f(x)$ para todo c e todo x então, escrevendo $a = f(1)$, tem-se $f(c) = f(c \cdot 1) = c \cdot f(1) = ca$, ou seja, $f(c) = ac$ para todo $c \in \mathbb{R}$. Numa notação mais adequada, temos $f(x) = ax$ para todo $x \in \mathbb{R}$, logo f é uma função linear.

Em suma, a definição tradicional equivale a dizer que a grandeza y é diretamente proporcional à grandeza x quando existe um número a (chamado a *constante de proporcionalidade*) tal que $y = ax$ para todo valor de x.

Quanto à proporcionalidade inversa, ela só tem sentido quando se trata de grandezas não-nulas. Seu modelo matemático é uma função $f\colon \mathbb{R}^* \to \mathbb{R}^*$ (onde $\mathbb{R}^* = \mathbb{R} - \{0\}$) tal que $f(cx) = f(x)/c$ para $c, x \in \mathbb{R}^*$ quaisquer. Usando o mesmo raciocínio anterior, isto quer

dizer que, para todo $x \in \mathbb{R}^*$, tem-se $f(x) = a/x$, onde a constante a é $f(1)$.

Fixaremos nossa atenção na proporcionalidade direta, que chamaremos apenas de "proporcionalidade".

Na prática, há situações em que a fórmula $y = ax$, que caracteriza a proporcionalidade, é dada explicitamente (ou quase). Por exemplo, se um quilo de açúcar custa a reais então x quilos custam $y = ax$ reais.

Em muitos casos, porém, a constante a de proporcionalidade não está clara e, às vezes, nem mesmo tem relevância alguma para o problema. Um exemplo disso se tem nas aplicações do teorema de Tales.

Naquele teorema, tem-se um triângulo ABC e uma correspondência que a cada ponto X do lado AB associa o ponto Y do lado AC tal que XY é paralelo a BC. O teorema de Tales assegura que o comprimento y do segmento AY é proporcional ao comprimento x de AX. Mas que importância tem a constante de proporcionalidade $a = y/x$? Por acaso, tem-se $a = \operatorname{sen} B / \operatorname{sen} C$ mas este valor não significa muito no caso.

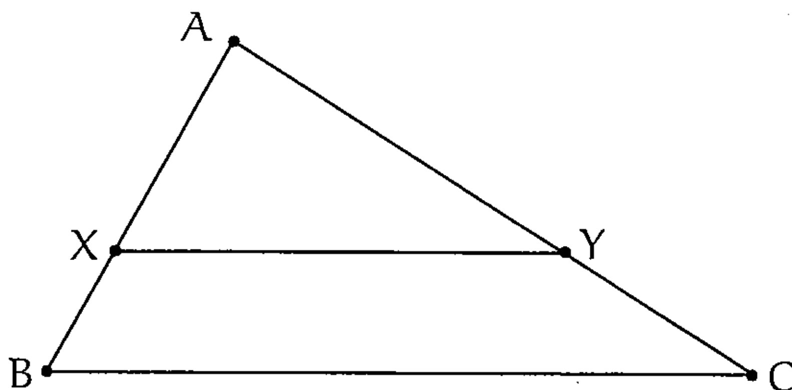


Figura 16

Este exemplo chama a atenção para o fato de que nos problemas relativos à proporcionalidade o que importa muitas vezes é saber apenas que se $y = f(x)$ e $y' = f(x')$ então $y'/x' = y/x$ é constante.

Quando a correspondência $x \mapsto y$, $x' \mapsto y'$ é uma proporcio-

nalidade, a igualdade $y'/x' = y/x$ permite que se determine um desses quatro números quando se conhecem os outros três. Nisto consiste a tradicional "regra de três".

Mas há uma questão preliminar que é a seguinte: como vamos ter certeza de que a correspondêcia $x \mapsto y$ é uma proporcionalidade? A definição de Trajano exige que se tenha $f(cx) = cf(x)$ para todos os valores reais de c e x. Em particular, *para todo* c. Isto é fácil de verificar quando c é inteiro. E nos outros casos? E se c for irracional? Felizmente basta que se saiba que $f(nx) = nf(x)$ para todo $x \in \mathbb{R}$ e todo n inteiro, desde que se suponha que f é monótona (o que é fácil de constatar na prática).

O teorema abaixo é a chave para determinar, em todas as situações, se uma dada função é ou não linear.

Teorema Fundamental da Proporcionalidade: *Seja* $f\colon \mathbb{R} \to \mathbb{R}$ *uma função crescente. As seguintes afirmações são equivalentes*:

(1) $f(nx) = nf(x)$ *para todo* $n \in \mathbb{Z}$ *e todo* $x \in \mathbb{R}$.

(2) *Pondo* $a = f(1)$, *tem-se* $f(x) = ax$ *para todo* $x \in \mathbb{R}$.

(3) $f(x+y) = f(x) + f(y)$ *para quaisquer* $x, y \in \mathbb{R}$.

Demonstração: Provaremos as implicações $(1) \Rightarrow (2)$, $(2) \Rightarrow (3)$ e $(3) \Rightarrow (1)$. A fim de mostrar que $(1) \Rightarrow (2)$, provemos inicialmente que, para todo número racional $r = m/n$, a hipótese (1) acarreta que $f(rx) = rf(x)$, seja qual for $x \in \mathbb{R}$. Com efeito, tem-se

$$n \cdot f(rx) = f(nrx) = f(mx) = m \cdot f(x),$$

logo

$$f(rx) = \frac{m}{n} \cdot f(x) = r \cdot f(x).$$

Seja $a = f(1)$. Como $f(0) = f(0 \cdot 0) = 0 \cdot f(0) = 0$, a monotonicidade de f nos dá $a = f(1) > f(0) = 0$. Assim, a é positivo. Além disso, temos $f(r) = f(r \cdot 1) = r \cdot f(1) = r \cdot a = ar$ para todo $r \in \mathbb{Q}$.

Mostremos agora que se tem $f(x) = ax$ para todo $x \in \mathbb{R}$.

Suponha, por absurdo, que exista algum número real x (necessariamente irracional) tal que $f(x) \neq ax$. Para fixar idéias,

admitamos $f(x) < ax$. (O caso $f(x) > ax$ seria tratado de modo análogo.) Temos

$$\frac{f(x)}{a} < x.$$

Tomemos um número racional r tal que

$$\frac{f(x)}{a} < r < x.$$

Então $f(x) < ar < ax$, ou seja, $f(x) < f(r) < ax$. Mas isto é absurdo, pois f é crescente logo, como $r < x$, deveríamos ter $f(r) < f(x)$. Esta contradição completa a prova de que $(1) \Rightarrow (2)$. As implicações $(2) \Rightarrow (3)$ e $(3) \Rightarrow (1)$ são óbvias.

Em algumas situações, o Teorema Fundamental da Proporcionalidade precisa ser aplicado a grandezas (como área ou massa, por exemplo) cujas medidas são expressas apenas por números positivos. Então temos uma função crescente $f\colon \mathbb{R}^+ \to \mathbb{R}^+$, onde $\mathbb{R}^+ = \{x \in \mathbb{R}; x > 0\}$ é o conjunto dos números positivos. Neste caso, as afirmações do Teorema lêem-se assim:

(1^+) $f(nx) = n \cdot f(x)$ para todo $n \in \mathbb{N}$ e todo $x \in \mathbb{R}^+$.

(2^+) Pondo $a = f(1)$, tem-se $f(x) = ax$ para todo $x \in \mathbb{R}^+$.

(3^+) $f(x+y) = f(x) + f(y)$ para quaisquer $x, y \in \mathbb{R}$.

Neste novo contexto, o Teorema Fundamental da Proporcionalidade continua válido, isto é, as afirmações (1^+), (2^+) e (3^+) são ainda equivalentes. Isto se mostra introduzindo a função $F\colon \mathbb{R} \to \mathbb{R}$, onde $F(0) = 0$, $F(x) = f(x)$ e $F(-x) = -f(x)$ para todo $x > 0$. Cada uma das afirmações (1^+), (2^+), (3^+) para f equivale a uma das afirmações (1), (2) e (3) para F.

Deve-se observar que a função f do teorema acima sendo crescente, tem-se $a = f(1) > 0$. No caso de se supor f decrescente vale um resultado análogo, com $a < 0$.

A importância deste teorema está no seguinte ponto: se queremos saber se $f\colon \mathbb{R} \to \mathbb{R}$ é uma função linear basta verificar duas coisas.

Primeira: f deve ser crescente ou decrescente. (Estamos deixando de lado o caso trivial de f identicamente nula.)

Segunda: $f(nx) = nf(x)$ para todo $x \in \mathbb{R}$ e todo $n \in \mathbb{Z}$. No caso de $f: \mathbb{R}^+ \to \mathbb{R}^+$ basta verificar esta última condição para $n \in \mathbb{N}$.

Exemplo 8. Se investirmos a quantia x, digamos numa caderneta de poupança, depois de um ano teremos um capital $f(x)$. Evidentemente, f é uma função crescente de x: quanto mais se aplica mais se recebe no final. Além disso, tem-se $f(nx) = nf(x)$ para todo $n \in \mathbb{N}$ e todo x. De fato, esta igualdade significa que tanto faz abrir uma caderneta de poupança com o capital inicial $x' = nx$ como abrir (no mesmo dia) n cadernetas, cada uma com o valor inicial x. O Teorema Fundamental nos permite concluir que $f(x)$ é proporcional a x. Mais precisamente, se a aplicação de 1 real der, no final de um ano, um valor de resgate igual a a, então o capital inicial de x reais se transformará em $f(x) = ax$ no final de um ano. (Não confundir este exemplo com o crescimento do capital em função do tempo. Este não é proporcional e será tratado quando estudarmos a função exponencial.)

Exemplo 9. Euclides dizia: "dois retângulos de mesma altura estão entre si como suas bases". Isto quer dizer que, se a altura de um retângulo é fixada, a área desse retângulo é proporcional à base. Ou ainda: a *área de um retângulo de altura* a *e base* x *é uma função linear de* x. É claro que esta afirmação é uma conseqüência super-óbvia da fórmula de área do retângulo. O ponto, todavia, é que ela é o argumento crucial para a dedução daquela fórmula, logo não pode ser deduzida como sua conseqüência. Para estabelecer sua veracidade, seja $f(x)$ a área do retângulo de altura a e base x. É claro que f é uma função crescente de x. Além disso, é claro que um retângulo de altura a e base nx pode ser decomposto em n retângulos de mesma altura a, cada um com base x, logo $f(nx) = nf(x)$. Segue-se, então, do teorema que $f(x) = A \cdot x$, onde $A = f(1)$ é a área de um retângulo de altura a e base 1. Vamos mostrar que $A = a$. O mesmo argumento, aplicado aos retângulos de mesma

base 1 e altura variável, mostra que $A = a \cdot U$, onde U é a área do retângulo de base e altura iguais a 1. Mas este é o quadrado de lado 1 o qual é, por definição, a unidade de área. Portanto $U = 1$ e $A = a$. Conclusão: a área de um retângulo de altura a e base x é igual a ax.

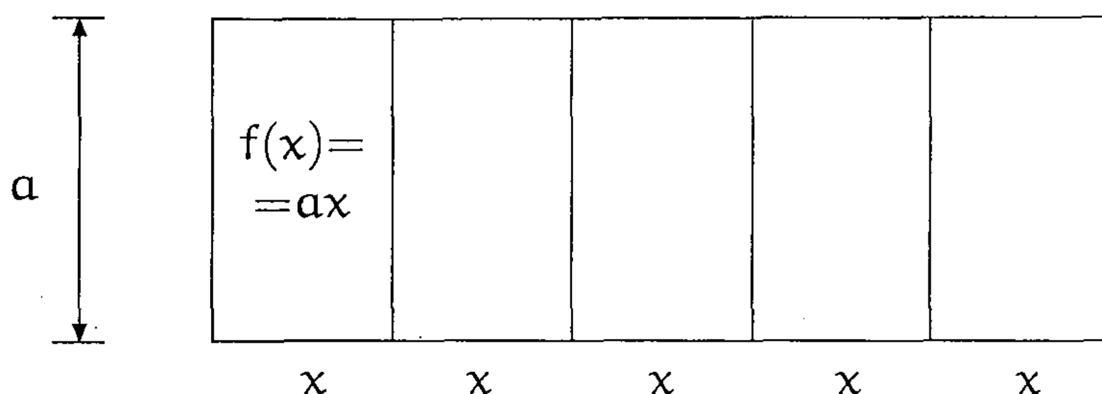


Figura 17

Observação: No enunciado que demos para o Teorema Fundamental da Proporcionalidade, fizemos a hipótese de que a função f fosse crescente (ou decrescente, seria o mesmo). Outra hipótese possível para o teorema – e equivalente, neste caso, à monotonicidade – seria de que a função f fosse contínua. Note-se que, na demonstração, a monotonicidade foi usada apenas para provar que se $f(r) = ar$ para todo r racional então $f(x) = ax$ para todo x real. Esta conclusão é imediata quando f é contínua, pois todo número real x é limite de uma seqüência de números racionais r_n, logo a continuidade de f nos dá $f(x) = \lim f(r_n) = \lim a\, r_n = ax$. A razão pela qual optamos em usar monotonicidade, em vez da continuidade para f é que este último conceito não é usualmente tratado no segundo grau, enquanto “crescente” e “decrescente” são noções bem mais elementares, que não dependem da idéia de limite.

4. Caracterização da Função Afim

Como saber se, numa determinada situação, o modelo matemático a ser adotado é uma função afim?

No caso da tarifa do taxi não há problema. Tem-se $f(x) = ax + b$ onde x é a distância percorrida, $f(x)$ é o preço a pagar, b é a bandeirada e a é a taxa por quilômetro rodado. Mas nem todo problema é assim tão explícito.

Vejamos um caso diferente.

E.W. observou, numa sapataria, que o vendedor determinava o número do sapato do cliente medindo seu pé com uma escala na qual, em vez de centímetros, estavam marcados os números $\ldots 36, 37, 38, \ldots$. O fato mais importante que ele percebeu foi que esses números estavam igualmente espaçados, isto é, a distância de cada um deles para o seguinte era constante. Isto queria dizer que a acréscimos iguais no tamanho do pé corresponderiam acréscimos iguais no número do sapato. Dito de outro modo: se um certo pé precisar de crescer h centímetros para passar de tamanho 33 para 34, precisará de crescer os mesmos h centímetros para passar de 38 para 39. Isto lhe deu a certeza de que a função que faz corresponder a cada comprimento x de um pé o número $f(x)$ do sapato adequado é uma função afim: $f(x) = ax + b$. (Vide teorema a seguir.)

E.W. sabia que, para determinar os coeficientes a, b da função afim, bastava conhecer $y_1 = f(x_1)$ e $y_2 = f(x_2)$ para dois valores diferentes quaisquer x_1 e x_2.

Ele atravessou a rua. Do outro lado havia uma papelaria, onde comprou uma régua. Voltou à sapataria e pediu emprestada a escala do vendedor. Como sua régua media até milímetros enquanto a escala só marcava pontos e meios pontos, escolheu dois valores $x_1 \neq x_2$ tais que os números de sapato correspondentes, $y_1 = f(x_1)$ e $y_2 = f(x_2)$, assinalados na escala, fossem inteiros. Tomou $x_1 = 20$, $x_2 = 28$ e viu que $f(x_1) = 32$, $f(x_2) = 42$. A partir daí, calculou os coeficientes $a = (y_1 - y_2)/(x_1 - x_2)$ e $b = y_1 - ax_1$ chegando à fórmula

$$f(x) = \frac{5x + 28}{4},$$

que dá o número do sapato de uma pessoa em função do compri-

mento do seu pé em centímetros. Para chegar à sua fórmula, E.W. fez uso do seguinte

Teorema: *Seja* $f: \mathbb{R} \to \mathbb{R}$ *uma função monótona injetiva. Se o acréscimo* $f(x+h) - f(x) = \varphi(h)$ *depender apenas de* h, *mas não de* x, *então* f *é uma função afim.*

A demonstração deste teorema, que faremos agora, é uma aplicação do Teorema Fundamental da Proporcionalidade. Para fixar idéias, suporemos que a função f seja crescente. Então $\varphi: \mathbb{R} \to \mathbb{R}$ também é crescente, com $\varphi(0) = 0$. Além disso, para quaisquer $h, k \in \mathbb{R}$ temos

$$\begin{aligned}\varphi(h+k) &= f(x+h+k) - f(x) \\ &= f((x+k)+h) - f(x+k) + f(x+k) - f(x) = \\ &= \varphi(h) + \varphi(k).\end{aligned}$$

Logo, pelo Teorema Fundamental da Proporcionalidade, pondo-se $a = \varphi(1)$, tem-se $\varphi(h) = a \cdot h$ para todo $h \in \mathbb{R}$. Isto quer dizer que $f(x+h) - f(x) = a \cdot h$. Chamando $f(0)$ de b, resulta $f(h) = ah + b$, ou seja, $f(x) = ax + b$ para todo $x \in \mathbb{R}$.

Observação. A recíproca do teorema acima é óbvia. Se $f(x) = ax + b$ então $f(x+h) - f(x) = ah$ não depende de x.

A hipótese de que $f(x+h) - f(x)$ não depende de x às vezes se exprime dizendo que "a acréscimos iguais de x correspondem acréscimos iguais para $f(x)$". Outra maneira de exprimir esta hipótese consiste em dizer que os acréscimos sofridos por $f(x)$ são proporcionais aos acréscimos dados a x.

Exemplo 10. Suponhamos um ponto que se movimenta sobre um eixo. Sua posição, em cada instante t, é determinada pela coordenada (abcissa) $f(t)$. Diz-se que se trata de um *movimento uniforme* quando o ponto se desloca sempre no mesmo sentido (isto é, f é uma função monótona) e, além disso, em tempos iguais percorre espaços iguais. Isto significa que $f(t+h)-f(t)$, espaço percorrido no tempo h, a partir da posição $f(t)$, depende apenas de h, mas não de

t. Então f é uma função afim: $f(t) = at+b$, onde $a = f(t+1)-f(t)$, espaço percorrido na unidade de tempo, chama-se a *velocidade* e $b = f(0)$ é a posição inicial.

Observação. Na definição usual de movimento uniforme, a condição de que o ponto móvel se desloque sempre no mesmo sentido não é imposta. A razão para isto é que se supõe sempre que, no movimento, a função $f(t)$ que dá a posição do ponto no instante t seja contínua. E, como já observamos antes, no Teorema Fundamental da Proporcionalidade, a monotonicidade de f pode ser substituída por sua continuidade, sem alterar a conclusão. Deve-se esclarecer, porém, que uma dessas hipóteses – monotonicidade, continuidade ou algo equivalente – deve ser incluída pois existem funções $f: \mathbb{R} \to \mathbb{R}$ incrivelmente complicadas, para as quais vale a condição $f(x+y) = f(x) + f(y)$ para $x, y \in \mathbb{R}$ quaisquer mas f não é da forma $f(x) = ax$.

Existe uma conexão interessante entre funções afins e progressões aritméticas, análoga à que veremos mais tarde entre funções exponenciais e progressões geométricas.

Uma *progressão aritmética* pode ser vista geometricamente como uma seqüência (finita ou infinita) de pontos $x_1, x_2, \ldots, x_i, \ldots$ igualmente espaçados na reta. Isto quer dizer que a *razão* $h = x_{i+1} - x_i$ não depende de i:

$$h = x_2 - x_1 = x_3 - x_2 = \cdots = x_{i+1} - x_i = \cdots .$$

Se $f: \mathbb{R} \to \mathbb{R}$ é uma função afim, digamos $f(x) = ax + b$, e $x_1, x_2, \ldots, x_i, \ldots$ é uma progressão aritmética, então os pontos $y_i = f(x_i)$, $i = 1, 2, \ldots$ também estão igualmente espaçados, isto é, formam uma progressão aritmética cuja razão é

$$y_{i+1} - y_i = (ax_{i+1} + b) - (ax_i + b) = a(x_{i+1} - x_i) = ah.$$

Assim, se tivermos uma reta não-vertical (gráfico de uma função afim) em $\mathbb{R}^2$ e tomarmos sobre ela os pontos

$$(1, y_1), (2, y_2), \ldots, (i, y_i), \ldots$$

cujas abcissas são os números naturais $1, 2, \ldots, i, \ldots$, as ordenadas $y_1, y_2, \ldots, y_i, \ldots$ desses pontos formam uma progressão aritmética.

Reciprocamente, se uma função monótona $f\colon \mathbb{R} \to \mathbb{R}$ transforma qualquer progressão aritmética $x_1, x_2, \ldots, x_i, \ldots$ numa progressão aritmética $y_1 = f(x_1), y_2 = f(x_2), \ldots, y_i = f(x_i), \ldots$ então f é uma função afim.

Com efeito, neste caso a nova função $g\colon \mathbb{R} \to \mathbb{R}$, definida por $g(x) = f(x) - f(0)$, transforma qualquer progressão aritmética noutra progressão aritmética, e agora tem a propriedade $g(0) = 0$. Mostremos que g é linear.

Para todo $x \in \mathbb{R}$, os números $-x$, 0, x formam uma progressão aritmética, logo o mesmo ocorre com os números $g(-x)$, 0, $g(x)$. Por conseguinte, $g(-x) = -g(x)$.

Em seguida, consideremos $x \in \mathbb{R}$ e $n \in \mathbb{N}$. Então os números $0, x, 2x, \ldots, nx$ formam uma progressão aritmética, o mesmo se dando com suas imagens por g: $0, g(x), g(2x), \ldots, g(nx)$. A razão desta progressão pode ser obtida tomando a diferença entre o segundo e o primeiro termo, logo esta razão é $g(x)$. Segue-se então que $g(nx) = n \cdot g(x)$. Finalmente, se n é um inteiro negativo, temos $-n \in \mathbb{N}$ logo $g(nx) = -g(-nx) = -(-n \cdot g(x)) = n \cdot g(x)$. Assim, vale $g(nx) = ng(x)$ para todo $n \in \mathbb{Z}$ e todo $x \in \mathbb{R}$. Pelo Teorema Fundamental da Proporcionalidade, segue-se que g é linear: $g(x) = ax$, portanto, pondo $f(0) = b$, temos

$$f(x) = g(x) + f(0) = ax + b$$

para todo $x \in \mathbb{R}$, como queríamos demonstrar.

5. Funções Poligonais

As funções poligonais surgem naturalmente, tanto na vida cotidiana (imposto de renda como função da renda líquida, preço de uma mercadoria que oferece descontos crescentes quando aumenta a quantidade comprada) como em diversas áreas da Matemática (Análise, Cálculo Numérico, Equações Diferenciais, To-

pologia).

Diz-se que $f\colon \mathbb{R} \to \mathbb{R}$ é uma *função poligonal* quando existem $t_0 < t_1 < \cdots < t_n$ tais que, para $x \leqslant t_0$, para $x \geqslant t_n$ e em cada um dos intervalos $[t_{i-1}, t_i]$, f coincide com uma função afim f_i. (Para evitar descontinuidades, exige-se que $f_i(t_i) = f_{i-1}(t_{i-1})$.) Equivalentemente, podemos dizer que uma função $f\colon \mathbb{R} \to \mathbb{R}$ é poligonal quando seu gráfico é uma linha poligonal.

O protótipo de função poligonal é uma função $f\colon \mathbb{R} \to \mathbb{R}$, definida por $f(x) = |x|$. Ou então $f(x) = |x - c|$, para algum $c \in \mathbb{R}$.

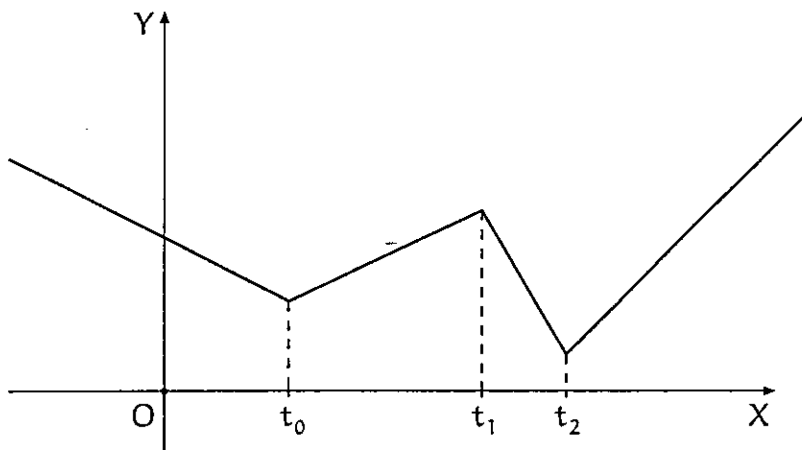


Figura 18

Outros exemplos são dados por expressões do tipo

$$f(x) = |\alpha x + \beta|$$

ou

$$g(x) = |x - \alpha| + |x - \beta|.$$

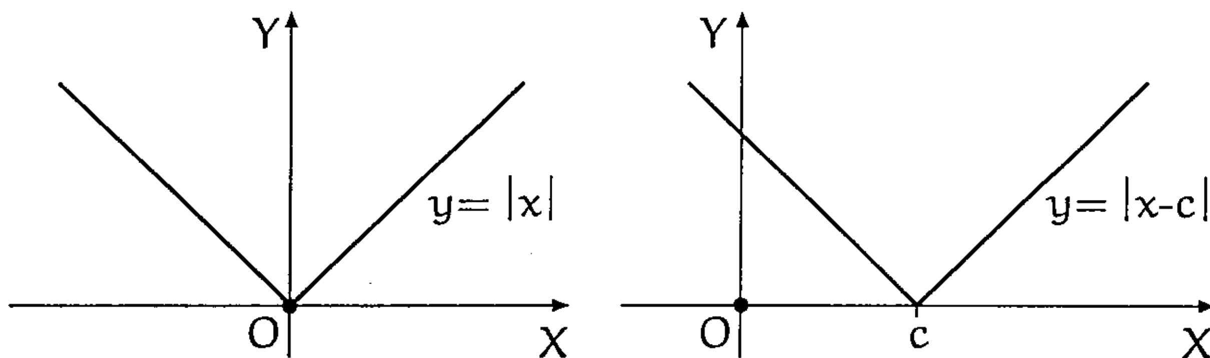


Figura 19

Estes exemplos nos levam a conjeturar que toda função poligonal pode ser definida combinando valores absolutos de funções afins. Esta conjetura é verdadeira. (Ver exercícios deste capítulo.)

Exercícios

1. Quando dobra o percurso em uma corrida de táxi, o custo da nova corrida é igual ao dobro, maior que o dobro ou menor que o dobro da corrida original?

2. A escala da figura abaixo é linear. Calcule o valor correspondente ao ponto assinalado.

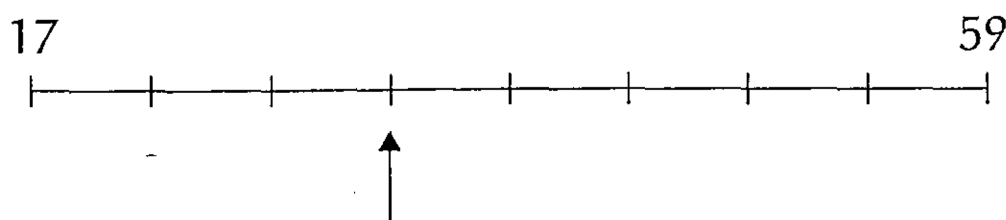


Figura 20

3. A escala N de temperaturas foi feita com base nas temperaturas máxima e mínima em Nova Iguaçu. A correspondência com a escala Celsius é a seguinte:

° N	° C
0	18
100	43

Em que temperatura ferve a água na escala N ?

4. Uma caixa dágua de 1000 litros tem um furo no fundo por onde escoa água a uma vazão constante. Ao meio dia de certo dia ela foi cheia e, às 6 da tarde desse dia, só tinha 850 litros. Quando ficará pela metade?

5. Um garoto brinca de arrumar palitos fazendo uma seqüência de quadrados como na figura Se ele fez n quadrados, quantos palitos utilizou?

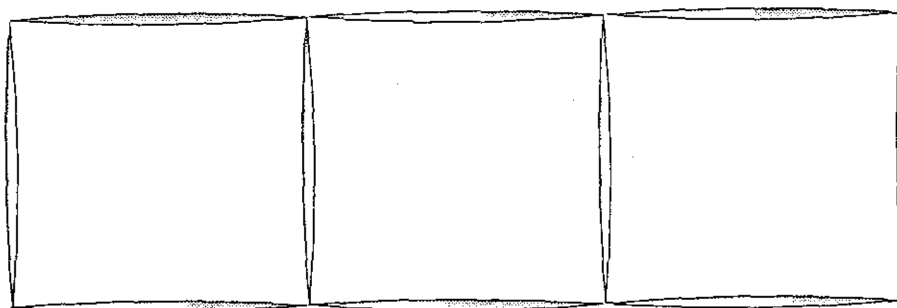


Figura 21

6. Admita que 3 operários, trabalhando 8 horas por dia, construam um muro de 36 metros em 5 dias.

a) Quantos dias são necessários para que uma equipe de 5 operários, trabalhando 6 horas por dia, construa um muro de 15 metros?

b) Que hipóteses foram implicitamente utilizadas na solução do item anterior?

c) Dentro dessas mesmas hipóteses, exprima o número D de dias necessários à construção de um muro em função do número N de operários, do comprimento C do muro e do número H de horas trabalhadas por dia.

7. As leis da Física, muitas vezes, descrevem relações de proporcionalidade direta ou inversa entre grandezas. Para cada uma das leis abaixo, escreva a expressão matemática correspondente.

a) (*Lei da gravitação universal*). Matéria atrai matéria na razão direta das massas e na razão inversa do quadrado das distâncias.

b) (*Gases perfeitos*). A pressão exercida por uma determinada massa de um gás é diretamente proporcional à temperatura absoluta e inversamente proporcional ao volume ocupado pelo gás.

c) (*Resistência elétrica*). A resistência de um fio condutor é diretamente proporcional ao seu comprimento e inversamente proporcional à área de sua seção reta.

d) (*Dilatação térmica*). A dilatação térmica sofrida por uma barra é diretamente proporcional ao comprimento da barra e à

variação de temperatura.

8. As grandezas X e Y são inversamente proporcionais. Se X sofre um acréscimo de 25% qual o decréscimo percentual sofrido por Y?

9. Os termos $a_1, a_2, \ldots, a_n$ de uma P.A. são os valores $f(1), f(2), \ldots, f(n)$ de uma função afim.

a) Mostre que cada a_i é igual à área de um trapézio delimitado pelo gráfico de f, pelo eixo OX e pelas retas verticais de equações

$$x = i - \frac{1}{2} \text{ e } x = i + \frac{1}{2}.$$

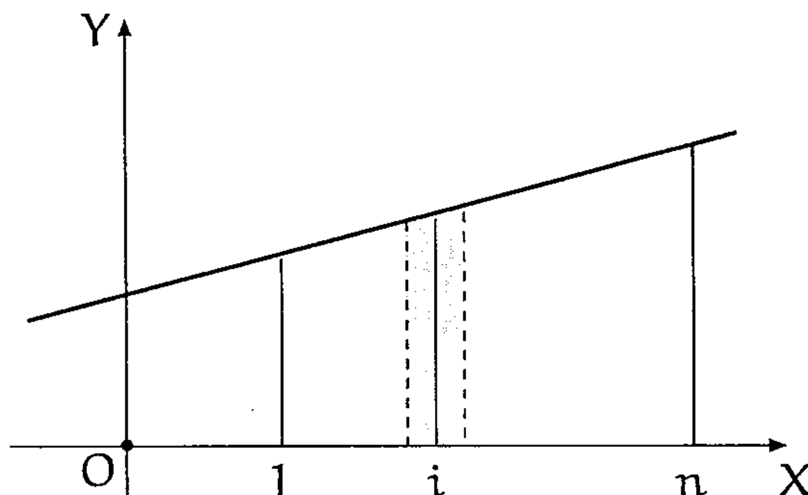


Figura 22

b) Mostre que a soma $S = a_1 + a_2 + \cdots + a_n$ é igual à área do trapézio delimitado pelo gráfico de f, pelo eixo OX e pelas retas verticais $x = \frac{1}{2}$ e $x = n + \frac{1}{2}$.

c) Conclua que $S = \dfrac{a_1 + a_n}{2} n$.

10. Pessoas apressadas podem diminuir o tempo gasto em uma escada rolante subindo alguns degraus da escada no percurso. Para uma certa escada, observa-se que uma pessoa gasta 30 segundos na escada quando sobe 5 degraus e 20 segundos quando sobe 10 degraus. Quantos são os degraus da escada e qual o tempo normalmente gasto no percurso?

11. Augusto, certo dia, fez compras em 5 lojas. Em cada loja, gastou metade do que possuia e pagou, na saida, R$ 2,00 de estacionamento. Se após toda essa atividade ainda ficou com R$ 20,00, que quantia ele tinha inicialmente?

12. Seguindo as idéias de E.W., construa uma régua para medir números de sapatos.

13. Estuda-se a implantação da chamada "fórmula 95". Por essa fórmula os trabalhadores teriam direito à aposentadoria quando a soma da idade com o número de anos de serviço atingisse 95. Adotada essa fórmula, quem começasse a trabalhar com 25 anos, com que idade se aposentaria?

14. Em uma escola há duas provas mensais, a primeira com peso 2 e a segunda com peso 3. Se o aluno não alcançar média 7 nessas provas, fará prova final. Sua média final será então a média entre a nota da prova final, com peso 2 e a média das provas mensais, com peso 3. João obteve 4 e 6 nas provas mensais. Se a média final para aprovação é 5, quanto ele precisa obter na prova final para ser aprovado?

15. Arnaldo dá a Beatriz tantos reais quanto Beatriz possui e dá a Carlos tantos reais quanto Carlos possui. Em seguida, Beatriz dá a Arnaldo e a Carlos tantos reais quanto cada um possui. Finalmente, Carlos faz o mesmo. Terminam todos com R$ 16,00 cada. Quanto cada um possuia no início?

16. Um carro sai de A para B e outro de B para A, simultaneamente, em linha reta, com velocidades constantes e se cruzam em um ponto situado a 720m do ponto de partida mais próximo. Completada a viagem, cada um deles pára por 10min e regressa, com a mesma velocidade da ida. Na volta, cruzam-se em um ponto situado a 400m do outro ponto de partida. Qual a distância de A até B ?

17. Em uma ferrovia, as estações A e B distam entre si 3 km e a cada 3 min parte um trem de cada uma delas em direção à outra.

Um pedestre parte de A para B, no exato momento em que um trem parte de A para B e outro chega a A vindo de B. Ele chega a B no exato momento em que um trem parte de B para A e outro trem chega a B vindo de A. Em seu caminho, o pedestre encontrou 17 trens que iam no mesmo sentido que ele e com 23 trens que iam no sentido oposto ao seu, aí incluídos os 4 trens já citados anteriormente. As velocidades dos trens são iguais. Calcule as velocidades dos trens e do pedestre.

18. Dado o gráfico da função f, abaixo, obtenha, em cada caso, o gráfico da função g tal que:

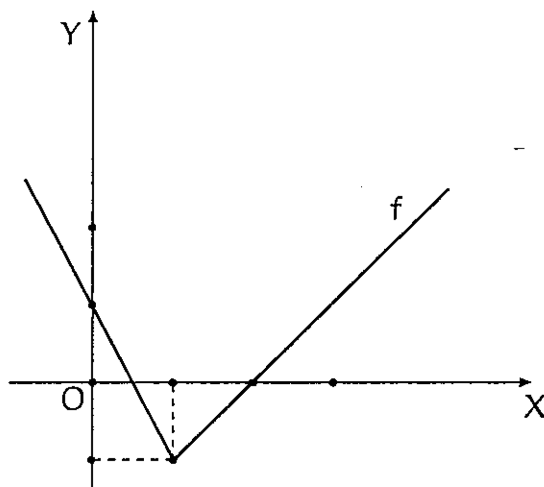


Figura 23

a) $g(x) = f(x) - 1$
b) $g(x) = f(x - 1)$
c) $g(x) = f(-x)$
d) $g(x) = 2(fx)$
e) $g(x) = f(2x)$
f) $g(x) = |f(x)|$
g) $g(x) = f(|x|)$
h) $g(x) = \max\{f(x), 0\}$

19. Determine os valores reais de x que satisfazem:
a) $2x + 3 - (x - 1) < x + 1$
b) $2x + 3 - (x - 1) < x + 5$

c) $\min\{x+1; 5-x\} > 2x-3$
d) $\min\{x+1; 5-x\} < 2x$
e) $\min\{2x-1; 6-x\} = x$
f) $2|x+1| - |1-x| \leqslant x+2$
g) $(2x+3)(1-x) = (2x+3)(x-2)$
h) $|x+1-|x-1|| \leqslant 2x-1$

20. Resolva a inequação

$$\frac{1}{2x+1} < \frac{1}{1-x}.$$

21. Determine a imagem da função $f: \mathbb{R} \to \mathbb{R}$ tal que $f(x) = \max\{x-1, 10-2x\}$.

22. Faça os gráficos de:
a) $f(x) = \min\{4-x; x+1\}$
b) $f(x) = |x+1| - |x-1|$

23. Identifique o conjunto dos pontos (x, y) tais que:
a) $|x| + |y| = 1$
b) $|x-y| = 1$

24. Um supermercado está fazendo uma promoção na venda de alcatra: um desconto de 10% é dado nas compras de 3 quilos ou mais. Sabendo que o preço do quilo de alcatra é de R$ 4,00, pede-se:
a) o gráfico do total pago em função da quantidade comprada.
b) o gráfico do preço médio por quilo em função da quantidade comprada.
c) a determinação de quais consumidores poderiam ter comprado mais alcatra pagando o mesmo preço.

25. Um supermercado está fazendo uma promoção na venda de alcatra: um desconto de 10% é dado nos quilos que excederem a 3. Sabendo que o preço do quilo de alcatra é de R$ 4,00, pede-se:
a) o gráfico do total pago em função da quantidade comprada.

b) o gráfico do preço médio por quilo em função da quantidade comprada.
c) a determinação de quantos quilos foram comprados por um consumidor que pagou R$ 15,00.

26. Os novos valores de IR-fonte:

Base de cálculo	**Alíquota**	**Parcela a deduzir**
Até R$ 900	Isento	–
De R$ 900 a R$ 1 800	15%	R$ 135
Acima de R$ 1 800	25%	R$ 315

Fonte: Secretaria da Receita Federal
Baseado na tabela acima, construa o gráfico do imposto a pagar em função do rendimento.

27. O imposto de renda y pago por uma pessoa que, em 1995, teve uma renda líquida y é calculado através de uma expressão da forma $y = ax - p$, onde a alíquota a e a parcela a deduzir p dependem da renda x e são dadas por uma tabela, parcialmente fornecida a seguir.

Renda (em R$)	**Alíquota** (a)	**Parcela a deduzir** (p)
Até 8800	0	0
De 8800 a 17160	15%	
De 17160 a 158450	26%	
Mais de 158450	35%	

a) Complete a tabela, de modo que o imposto a pagar varie continuamente com a renda (isto é, não haja saltos ao se passar de uma faixa de renda para outra).
b) Se uma pessoa está na terceira faixa e sua renda aumenta de R$ 5.000,00, qual será seu imposto adicional (supondo que este acréscimo não acarrete uma mudança de faixa)?

c) É comum encontrar pessoas que lamentam estar no início de uma faixa de taxação ("que azar ter recebido este dinheiro a mais!"). Este tipo de reclamação é procedente?

d) Os casais têm a alternativa de apresentar declaração em conjunto ou separadamente. No primeiro caso, o "cabeça do casal" pode efetuar uma dedução de R$ 3.000,00 em sua renda líquida mas, em compensação, tem que acrescentar a renda do cônjuge. Em que casos é vantajosa a declaração em separado?

e) A tabela de taxação é, às vezes, dada de uma outra forma, para permitir o cálculo do imposto através de uma expressão da forma $y = b(x - q)$ (isto é, primeiro se deduz a parcela q e depois se aplica a alíquota). Converta a tabela acima para este formato (isto é, calcule os valores de b e q para cada faixa de renda).

f) Qual a renda para a qual o imposto é igual a R$ 20.000,00?

g) Esboce o gráfico da função que associa a cada renda x o percentual desta renda que é pago de imposto.

28. Uma copiadora publicou a seguinte tabela de preços:

Número de cópias de um mesmo original	Preço por cópia
de 1 a 19	R$ 0,10
de 20 a 49	R$ 0,08
50 ou mais	R$ 0,06

a) Esboce o gráfico da função que associa a cada natural n o custo de n cópias de um mesmo original.

b) O uso da tabela acima provoca distorções. Aponte-as e sugira uma tabela de preços mais razoável.

29. Discutir o número de soluções da equação $|x - 2| = ax + b$ e, função dos parâmetros a e b.

30. Chama-se de *função rampa* a uma função poligonal $f: [a, b] \to \mathbb{R}$, cujo gráfico é de uma das formas abaixo:

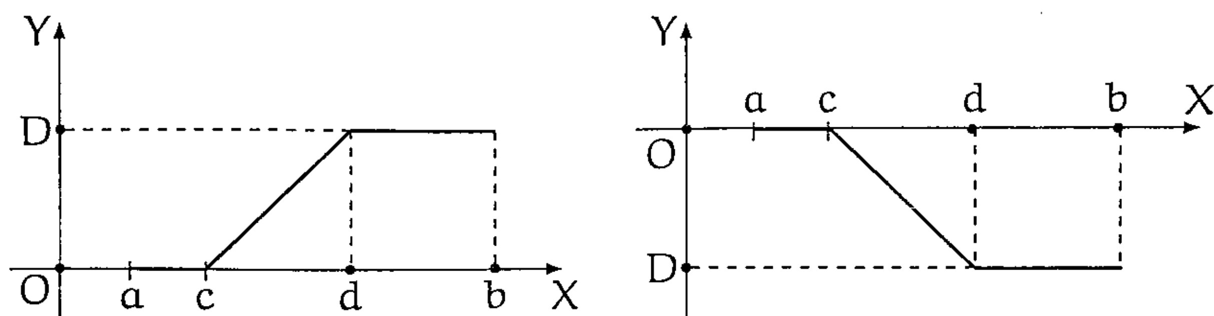


Figura 23

Isto é, f tem dois patamares $[a, c]$ e $[d, b]$, onde assume, respectivamente, os valores 0 e D, ligados por uma rampa.

a) Mostre que toda função rampa pode ser escrita na forma

$$f(x) = \frac{\alpha}{2}[(d-c) + |x-c| - |x-d|],$$

para todo $x \in [a, b]$, onde

$$\alpha = \frac{D}{d-c}$$

é a inclinação da rampa.

b) Mostre que toda função poligonal definida em um intervalo $[a, b]$ pode ser expressa como uma soma de uma função constante (que pode ser vista como uma função rampa de inclinação zero) com um número finito de funções rampa. Escreva nesta forma a função poligonal cujo gráfico é dado abaixo.

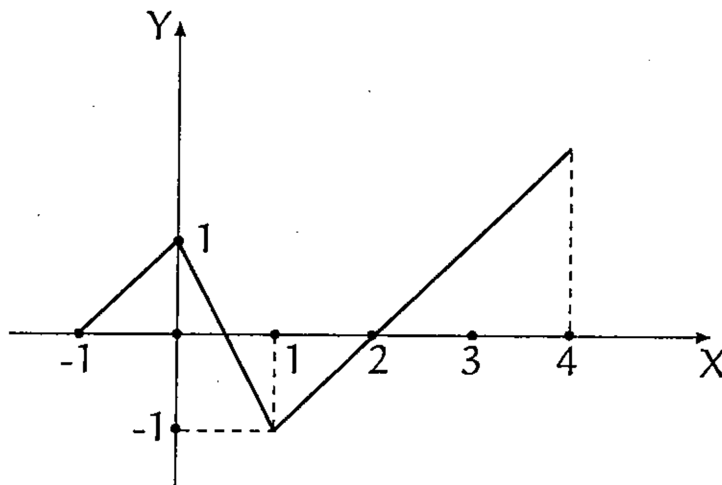


Figura 24

c) Conclua que toda função poligonal definida em um intervalo $[a, b]$ pode ser escrita na forma

$$f(x) = A + \alpha_1|x - a_1| + \alpha_2|x - a_2| + \cdots + \alpha_n|x - a_n|,$$

para todo $x \in [a, b]$, onde $a_1, a_2, \ldots, a_n$ são as abcissas dos vértices da poligonal. Escreva nesta forma a função poligonal cujo gráfico é dado acima.

31. Dadas as progressões aritméticas

$$(a_1, a_2, \ldots, a_n, \ldots) \quad \text{e} \quad (b_1, b_2, \ldots, b_n, \ldots),$$

mostre que existe uma, e somente uma, função afim $f\colon \mathbb{R} \to \mathbb{R}$ tal que $f(a_1) = b_1, f(a_2) = b_2, \ldots, f(a_n) = b_n, \ldots$

32. A e B são locadoras de automóvel. A cobra 1 real por quilômetro rodado mais uma taxa de 100 reais fixa. B cobra 80 centavos por quilômetro mais uma taxa fixa de 200 reais. Discuta a vantagem de A sobre B ou de B sobre A em função do número de quilômetros a serem rodados.

33. Defina uma função $f\colon \mathbb{R} \to \mathbb{R}$ pondo $f(x) = 2x$ se x é racional e $f(x) = 3x$ se x é irracional. Mostre que se tem $f(nx) = nf(x)$ para todo $n \in \mathbb{Z}$ e todo $x \in \mathbb{R}$ mas f não é linear.

34. Prove que a função $f\colon \mathbb{R} \to \mathbb{R}$, definida por $f(x) = 3x + \operatorname{sen}(2\pi x)$, é crescente e, para todo $x \in \mathbb{R}$ fixado, transforma a progressão aritmética $x, x+1, x+2, \ldots$ numa progressão aritmética. Entretanto, f não é afim. Por que isto não contradiz o fato provado no final da seção 4 (pág. 102)?

Capítulo 6

Funções Quadráticas

1. Definição e Preliminares

Uma função $f: \mathbb{R} \to \mathbb{R}$ chama-se *quadrática* quando existem números reais a, b, c, com $a \neq 0$, tais que $f(x) = ax^2 + bx + c$ para todo $x \in \mathbb{R}$.

A primeira observação que faremos é: os coeficientes a, b, c da função quadrática f ficam inteiramente determinados pelos valores que essa função assume. Noutras palavras, se $ax^2 + bx + c = a'x^2 + b'x + c'$ para todo $x \in \mathbb{R}$ então $a = a'$, $b = b'$ e $c = c'$.

Com efeito, seja $ax^2 + bx + c = a'x^2 + b'x + c'$ para todo $x \in \mathbb{R}$. Tomando $x = 0$, obtemos $c = c'$. Então, cortando c e c', tem-se $ax^2 + bx = a'x^2 + b'x$ para todo $x \in \mathbb{R}$. Em particular, esta igualdade vale para todo $x \neq 0$. Neste caso, cancelando x, obtemos $ax + b = a'x + b'$ para todo $x \neq 0$. Fazendo primeiro $x = 1$ e depois $x = -1$, vem $a + b = a' + b'$ e $-a + b = -a' + b'$, donde concluímos $a = b$ e $a' = b'$.

A observação acima permite que se identifique uma função quadrática com um trinômio do segundo grau. Há, em princípio, uma diferença sutil entre esses dois conceitos, Um *trinômio do segundo grau* é uma expressão formal do tipo $aX^2 + bX + c$, com $a, b, c \in \mathbb{R}$, sendo $a \neq 0$. A palavra *formal* aí significa que a letra X é apenas um símbolo, sendo X^2 um outro modo de escrever XX. Por definição, dois trinômios $aX^2 + bX + c$ e $a'X^2 + b'X + c'$ são iguais quando $a = a'$, $b = b'$ e $c = c'$. [Em última análise, um trinômio é o mesmo que um terno ordenado de números reais (a, b, c).]

A cada trinômio corresponde a função quadrática definida por $x \mapsto ax^2 + bx + c$. A observação anterior significa que essa correspondência (trinômio) $\mapsto$ (função quadrática) é biunívoca. (Pela definição de função quadrática, tal correspondência é automaticamente sobrejetiva.)

Exemplo 1. As frações racionais

$$\frac{X^3 - 3X + 2}{X^2 - 2X + 1} \quad \text{e} \quad \frac{X^4 + X^3 - X^2 + X - 2}{X^3 - X^2 + X - 1}$$

são expressões formalmente bem diferentes, que definem a mesma função $f\colon \mathbb{R} - \{1\} \to \mathbb{R}$, pois, para todo número real $x \neq 1$, tem-se

$$\frac{x^3 - 3x + 2}{x^2 - 2x + 1} = \frac{x^4 + x^3 - x^2 + x - 2}{x^3 - x^2 + x - 1} = x + 2.$$

Este exemplo serve para mostrar que, quando não se trata de polinômios, duas expressões formais distintas podem definir a mesma função real de uma variável real.

A partir de agora, identificaremos a função quadrática com o trinômio do segundo grau a ela associado e nos permitiremos falar da função

$$f(x) = ax^2 + bx + c$$

sempre que não houver perigo de confundí-la com o número real $f(x)$, que é o valor por ela assumido no ponto x.

A fim de que se tenha $a = a'$, $b = b'$ e $c = c'$, não é necessário exigir, como fizemos acima, que

$$ax^2 + bx + c = a'x^2 + b'x + c'$$

para todo $x \in \mathbb{R}$. Basta supor que esta igualdade valha para três valores distintos de x. Passemos a discutir este assunto.

Suponhamos que as funções quadráticas

$$f(x) = ax^2 + bx + c \quad \text{e} \quad g(x) = a'x^2 + b'x + c'$$

assumam os mesmos valores $f(x_1) = g(x_1)$, $f(x_2) = g(x_2)$ e $f(x_3) = g(x_3)$ para três números reais distintos x_1, x_2 e x_3. Escrevendo

$\alpha = a - a'$, $\beta = b - b'$ e $\gamma = c - c'$, queremos mostrar que $\alpha = \beta = \gamma = 0$. Sabemos que $f(x_1) - g(x_1) = 0$, $f(x_2) - g(x_2) = 0$ e $f(x_3) - g(x_3) = 0$. Isto significa que

$$\text{(S)} \qquad \begin{cases} \alpha x_1^2 + \beta x_1 + \gamma &= 0 \\ \alpha x_2^2 + \beta x_2 + \gamma &= 0 \\ \alpha x_3^2 + \beta x_3 + \gamma &= 0 \end{cases}$$

Subtraindo a primeira equação de cada uma das outras, vem:

$$\alpha(x_2^2 - x_1^2) + \beta(x_2 - x_1) = 0$$
$$\alpha(x_3^2 - x_1^2) + \beta(x_3 - x_1) = 0$$

Como $x_2 - x_1 \neq 0$ e $x_3 - x_1 \neq 0$, podemos dividir a primeira destas equações por $x_2 - x_1$ e a segunda por $x_3 - x_1$, obtendo

$$\alpha(x_1 + x_2) + \beta = 0$$
$$\alpha(x_1 + x_3) + \beta = 0$$

Subtraindo membro a membro, temos $\alpha(x_3 - x_2) = 0$.

Como $x_3 - x_2 \neq 0$, resulta daí que $\alpha = 0$. Substituindo nas equações anteriores, obtemos sucessivamente $\beta = 0$ e $\gamma = 0$.

Acabamos de mostrar que *se duas funções quadráticas assumem os mesmos valores em três pontos distintos* x_1, x_2, x_3 *então essas funções são iguais, isto é, assumem o mesmo valor para qualquer número real* x.

Examinando o argumento usado, vemos que se tem um sistema (S) de três equações lineares a três incógnitas α, β, γ com os segundos membros iguais a zero (sistema homogêneo). O que provamos foi que a única solução desse sistema é a solução trivial $\alpha = \beta = \gamma = 0$. Sabemos que, em geral, quando um sistema homogêneo só admite a solução trivial então podemos substituir os zeros dos segundos membros por números arbitrários que sempre teremos solução única. No caso presente, isto é fácil de ver diretamente: usando os mesmos passos seguidos acima, vemos que, dados arbitrariamente os números reais y_1, y_2, y_3, existe um, e

somente um terno ordenado de número a, b, c tais que

$$ax_1^2 + bx_1 + c = y_1$$
$$ax_2^2 + bx_2 + c = y_2$$
$$ax_3^2 + bx_3 + c = y_3\,.$$

Neste sistema, vários hábitos tradicionais são violados. As incógnitas são a, b, c em vez dos x, y, z de costume. Os coeficientes conhecidos são x_1, x_2, $x_3 x_1^2$, x_2^2, x_3^2 e 1, 1, 1. Além disso, as incógnitas estão escritas antes dos coeficientes. Mesmo assim, não há maiores dificuldades em resolvê-lo, adotando, como dissemos, a mesma seqüência de passos do caso homogêneo.

Estamos especialmente interessados no valor da incógnita a neste sistema. Ele é

$$a = \frac{1}{x_3 - x_2}\left[\frac{y_3 - y_1}{x_3 - x_1} - \frac{y_2 - y_1}{x_2 - x_1}\right].$$

Podemos então afirmar o seguinte: dados três números reais distintos x_1, x_2, x_3 e números reais arbitrários y_1, y_2, y_3, existe um, e somente um, terno de números a, b, c tais que a função

$$f(x) = ax^2 + bx + c$$

cumpre $f(x_1) = y_1$, $f(x_2) = y_2$ e $f(x_3) = y_3$.

A função $f(x) = ax^2 + bx + c$, acima obtida, pode não ser quadrática, a menos que nos asseguremos que $a \neq 0$. O valor de a acima obtido mostra que a é zero se, e somente se, vale

$$\frac{y_3 - y_1}{x_3 - x_1} = \frac{y_2 - y_1}{x_2 - x_1}.$$

Se olharmos para os pontos $A = (x_1, y_1)$, $B = (x_2, y_2)$ e $C = (x_3, y_3)$ em $\mathbb{R}^2$, a condição acima significa que as retas AC e AB têm a mesma inclinação, isto é, que os pontos A, B e C são colineares.

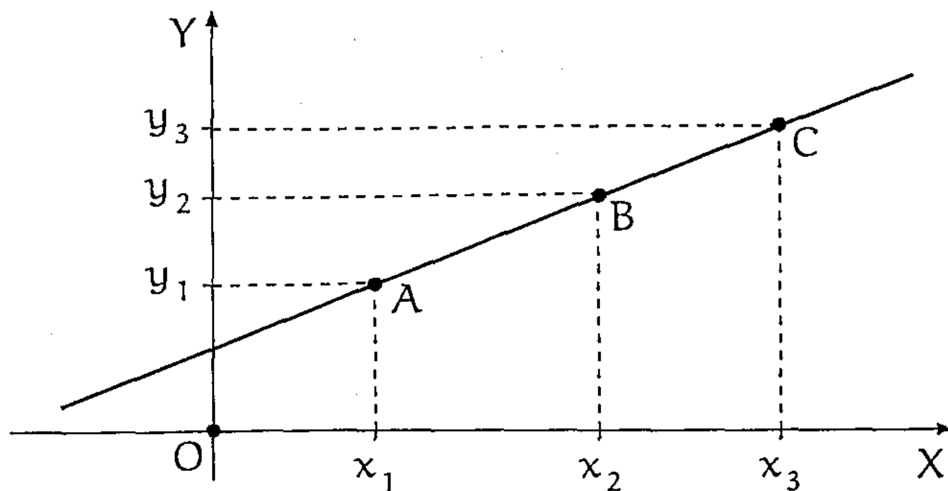


Figura 25

Então podemos enunciar:

Sejam x_1, x_2, x_3 *três números reais distintos e* y_1, y_2, y_3 *números tais que os pontos* $A = (x_1, y_1)$, $B = (x_2, y_2)$ *e* $C = (x_3, y_3)$ *são não-colineares em* $\mathbb{R}^2$. *Existe uma, e somente uma, função quadrática* $f(x) = ax^2 + bx + c$ *tal que* $f(x_1) = y_1$, $f(x_2) = y_2$ *e* $f(x_3) = y_3$.

Comentário sobre Colinearidade

Sejam $A = (x_1, y_1)$, $B = (x_2, y_2)$ e $C = (x_3, y_3)$ três pontos distintos em $\mathbb{R}^2$. A condição necessária e suficiente para que esses pontos sejam colineares é apresentada, em todos os nossos textos escolares, sob a forma da equação

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0,$$

na qual o primeiro membro é um determinante 3×3. Desenvolvendo esse determinante, vemos que a equação acima significa

$$(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1) = 0,$$

ou seja

$$(*) \qquad \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_3 - y_1}{x_3 - x_1},$$

Como vimos, esta última igualdade exprime que as retas AB e AC têm a mesma inclinação. Ela constitui um critério de colinearidade mais simples, mais direto, mais fácil de verificar e mais elementar do que aquele adotado nos livros que nossos alunos usam, pois não requer o conhecimento de determinantes.

Pode-se objetar que a igualdade (*) só tem sentido quando $x_1 \neq x_2$ e $x_1 \neq x_3$. É verdade. Mas o caso em que $x_1 = x_2$ ou $x_1 = x_3$ não requer cálculo algum. Se algum dos denominadores na igualdade (*) é igual a zero, isto quer dizer que dois dos pontos A, B, C têm a mesma abcissa, logo estão sobre uma reta vertical. Basta então olhar para a abcissa do terceiro ponto: se for igual às outras duas então A, B e C estão na mesma vertical, logo são colineares. Se for diferente, A, B e C não são colineares.

2. Um Problema Muito Antigo

O estudo das funções quadráticas tem sua origem na resolução da equação do segundo grau.

Problemas que recaem numa equação do segundo grau estão entre os mais antigos da Matemática. Em textos cuneiformes, escritos pelos babilônios há quase quatro mil anos, encontramos, por exemplo, a questão de achar dois números conhecendo sua soma s e seu produto p.

Em termos geométricos, este problema pede que se determinem os lados de um retângulo conhecendo o semi-perímetro s e a área p.

Os números procurados são as raízes da equação do segundo grau

$$x^2 - sx + p = 0.$$

Com efeito, se um dos números é x, o outro é $s-x$ e seu produto é

$$p = x(s-x) = sx - x^2,$$

logo

$$x^2 - sx + p = 0.$$

Observe que se α é uma raiz desta equação, isto é, $\alpha^2 - s\alpha + p = 0$, então $\beta = s - \alpha$ também é raiz, pois

$$\begin{aligned}\beta^2 - s\beta + p &= (s-\alpha)^2 - s(s-\alpha) + p = \\ &= s^2 - 2s\alpha + \alpha^2 - s^2 + s\alpha + p = \\ &= \alpha^2 - s\alpha + p = 0.\end{aligned}$$

Achar as raízes da equação $x^2 - sx + p = 0$ é, também, um conhecimento milenar. Note-se que, até o fim do século 16, não se usava uma fórmula para os valores das raízes, simplesmente porque não se representavam por letras os coeficientes de uma equação. Isto começou a ser feito a partir de François Viète, matemático francês que viveu de 1540 a 1603. Antes disso, o que se tinha era uma receita que ensinava como proceder em exemplos concretos (com coeficientes numéricos).

A regra para achar dois números cuja soma e cujo produto são dados era assim enunciada pelos babilônios:

Eleve ao quadrado a metade da soma, subtraia o produto e extraia a raiz quadrada da diferença. Some ao resultado a metade da soma. Isso dará o maior dos números procurados. Subtraia-o da soma para obter o outro número.

Na notação atual, esta regra fornece as raízes

$$x = \frac{s}{2} + \sqrt{\left(\frac{s}{2}\right)^2 - p} \quad \text{e} \quad s - x = \frac{s}{2} - \sqrt{\left(\frac{s}{2}\right)^2 - p}$$

para a equação $x^2 - sx + p = 0$.

Os autores dos textos cuneiformes não deixaram registrado o argumento que os levou a esta conclusão, mas há indícios de que pode ter sido algo assim:

Sejam α e β os números procurados, digamos com $\alpha \leqslant \beta$. Esses números α e β são equidistantes da média aritmética $\frac{s}{2} = \frac{\alpha+\beta}{2}$. Se conhecermos a diferença $d = \beta - (s/2) = (s/2) - \alpha$ teremos os dois números $\alpha = (s/2) - d$ e $\beta = (s/2) + d$. Mas d é fácil de achar,

pois

$$p = \alpha\beta = \left(\frac{s}{2} - d\right)\left(\frac{s}{2} + d\right) = \left(\frac{s}{2}\right)^2 - d^2,$$

logo

$$d^2 = \left(\frac{s}{2}\right)^2 - p \quad \text{e} \quad d = \sqrt{\left(\frac{s}{2}\right)^2 - p}.$$

Daí

$$\alpha = \frac{s}{2} - d = \frac{s}{2} - \sqrt{\left(\frac{s}{2}\right)^2 - p}$$

e

$$\beta = \frac{s}{2} + d = \frac{s}{2} + \sqrt{\left(\frac{s}{2}\right)^2 - p}.$$

Como os dados s e p do problema eram sempre números positivos, os babilônios nunca tiveram preocupação com eventuais soluções negativas fornecidas por sua regra. Mas certamente deviam ocorrer casos em que $(s/2)^2 < p$, como no problema de achar dois números cuja soma e cujo produto são ambos iguais a 2. Isto porém não os levou a inventarem os números complexos. Nestes casos, eles simplesmente diziam que os números procurados não existiam. O que é absolutamente correto no âmbito dos números reais.

Observação 1. Os números complexos só vieram a forçar sua admissão na Matemática no século 16, com a fórmula para as raízes da equação do terceiro grau, que fornecia as raízes reais por meio de uma expressão contendo raízes quadradas de números negativos.

Observação 2. Se procurarmos dois números cuja soma é 6 e cujo produto é 9, encontraremos que esses números são 3 e 3. Então é um número só; não são dois. Para não ter que acrescentar ao enunciado do nosso problema a frase "... ou um número cujo dobro é s e cujo quadrado é p," preferimos seguir o costume, que se adota em Matemática desde aqueles tempos, segundo o qual

a palavra "dois" às vezes significa "dois ou um". Quando quisermos garantir que significa "dois" mesmo, diremos "dois números diferentes". Mesma observação vale para três, quatro, etc.

3. A Forma Canônica do Trinômio

Consideremos o trinômio

$$ax^2 + bx + c = a\left[x^2 + \frac{b}{a}x + \frac{c}{a}\right].$$

As duas primeiras parcelas dentro do colchete são as mesmas do desenvolvimento do quadrado $\left(x + \frac{b}{2a}\right)^2$. Completando o quadrado, podemos escrever:

$$ax^2 + bx + c = a\left[x^2 + 2 \cdot \frac{b}{2a} \cdot x + \frac{b^2}{4a^2} - \frac{b^2}{4a^2} + \frac{c}{a}\right],$$

ou:

$$ax^2 + bx + c = a\left[\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a^2}\right].$$

Esta maneira de escrever o trinômio do segundo grau (chamada a *forma canônica*) tem algumas conseqüências.

Em primeiro lugar, ela conduz imediatamente à fórmula que dá as raízes da equação $ax^2 + bx + c = 0$. Com efeito, sendo $a \neq 0$, temos as seguintes equivalências

$$ax^2 + bx + c = 0 \Leftrightarrow \left(x + \tfrac{b}{2a}\right)^2 + \tfrac{4ac - b^2}{4a^2} = 0 \quad (1)$$

$$\Leftrightarrow \left(x + \tfrac{b}{2a}\right)^2 = \tfrac{b^2 - 4ac}{4a^2} \quad (2)$$

$$\Leftrightarrow x + \tfrac{b}{2a} = \pm\tfrac{\sqrt{b^2 - 4ac}}{2a} \quad (3)$$

$$\Leftrightarrow x = \tfrac{-b + \sqrt{b^2 - 4ac}}{2a}. \quad (4)$$

A passagem da linha (2) para a linha (3) só tem sentido quando o *discriminante*

$$\Delta = b^2 - 4ac$$

é $\geqslant 0$. Caso tenhamos $\Delta < 0$, a equivalência entre as linhas (1) e

(2) significa que a equação dada não possui solução real, pois o quadrado de $x + (b/2a)$ não pode ser negativo.

O método de completar o quadrado tem aplicações noutras questões matemáticas. Independente disso, é instrutivo fazer os alunos praticarem seu uso em exemplos concretos, para resolverem a equação do segundo grau sem aplicar diretamente a fórmula (4).

Da fórmula (4) resulta imediatamente que, se o discriminante $\Delta = b^2 - 4ac$ é positivo, a equação

$$ax^2 + bx + c = 0$$

tem duas raízes reais distintas

$$\alpha = (-b - \sqrt{\Delta})/2a$$

e

$$\beta = (-b + \sqrt{\Delta})/2a,$$

com $\alpha < \beta$, cuja soma é $s = -b/a$ e cujo produto é

$$p = (b^2 - \Delta)/4a^2 = 4ac/4a^2 = c/a.$$

Em particular, a média aritmética das raízes é $-b/2a$, ou seja, as raízes α e β são equidistantes do ponto $-b/2a$.

Quando $\Delta = 0$, a equação dada possui uma única raiz, chamada *raiz dupla*, igual a $-b/2a$.

Suponhamos $a > 0$. A forma canônica

$$f(x) = ax^2 + bx + c = a\left[\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a^2}\right]$$

exibe, no interior dos colchetes, uma soma de duas parcelas. A primeira depende de x e é sempre $\geqslant 0$. A segunda é constante. O menor valor dessa soma é atingido quando

$$\left(x + \frac{b}{2a}\right)^2$$

é igual a zero, ou seja, quando $x = -b/2a$. Neste ponto, $f(x)$ também assume seu valor mínimo. Portanto, quando $a > 0$, o menor valor assumido por

$$f(x) = ax^2 + bx + c$$

é

$$f(-b/2a) = c - (b^2/4a).$$

Se $a < 0$, o valor $f(-b/2a)$ é o maior dos números $f(x)$, para qualquer $x \in \mathbb{R}$.

Quando $a > 0$, $f(x) = ax^2 + bx + c$ não assume valor máximo: é uma função ilimitada superiormente. Analogamente, quando $a < 0$, $f(x)$ não assume valor mínimo: é ilimitada inferiormente.

A forma canônica ainda nos ajuda a responder a seguinte pergunta: Dada a função quadrática $f(x) = ax^2 + bx + c$, para quais valores $x \neq x'$ tem-se $f(x) = f(x')$?

Olhando para a forma canônica, vemos que $f(x) = f(x')$ se, e somente se,

$$\left(x + \frac{b}{2a}\right)^2 = \left(x' + \frac{b}{2a}\right)^2.$$

Como estamos supondo $x \neq x'$, isto significa que

$$x' + \frac{b}{2a} = -\left(x + \frac{b}{2a}\right),$$

isto é

$$\frac{x + x'}{2} = -\frac{b}{2a}.$$

Portanto, a função quadrática $f(x) = ax^2 + bx + c$ assume o mesmo valor $f(x) = f(x')$ para $x \neq x'$ se, e somente se, os pontos x e x' são equidistantes de $-b/2a$.

Exemplo 2. O conhecimento do ponto onde uma função quadrática assume seu valor máximo ou mínimo permite obter rapidamente uma resposta para a tradicional questão de saber qual o valor máximo do produto de dois números cuja soma é constante. Neste problema, um número s é dado e quer-se achar um par de

números x, y, com $x + y = s$, tais que o produto xy seja o maior possível. De $x+y = s$ tiramos $y = s-x$, portanto deve-se encontrar o valor de x que torna máximo o produto $x(s-x) = -x^2 + sx$. Esse valor máximo é assumido quando $x = s/2$, logo $y = s - x = s/2$. Concluímos então que o produto de dois números cuja soma é constante assume seu valor máximo quando esses números são iguais. (Note como ficaria complicado o enunciado desta conclusão se não tivéssemos permitido, em alguns casos, que dois seja igual a um.)

4. O Gráfico da Função Quadrática

O gráfico de uma função quadrática é uma parábola.

Dados um ponto F e uma reta d que não o contém, a *parábola* de *foco* F e *diretriz* d é o conjunto dos pontos do plano que distam igualmente de F e de d.

A reta perpendicular à diretriz, baixada a partir do foco, chama-se o *eixo* da parábola. O ponto da parábola mais próximo da diretriz chama-se o *vértice* dessa parábola. Ele é o ponto médio do segmento cujas extremidades são o foco e a interseção do eixo com a diretriz.

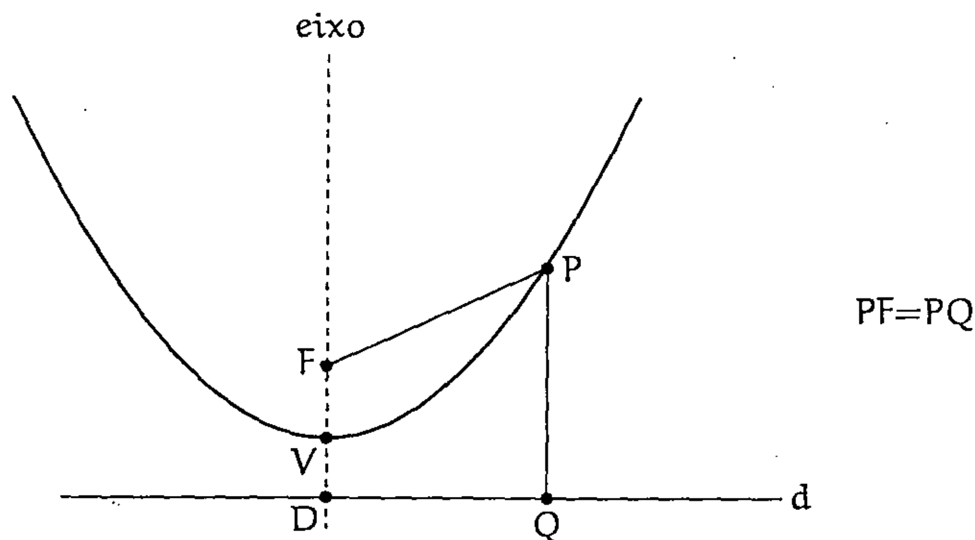


Figura 26

Lembremos que a distância de um ponto a uma reta é o comprimento do segmento perpendicular baixado do ponto sobre a reta.

Exemplo 3. O gráfico da função quadrática $f(x) = x^2$ é a parábola cujo foco é $F = (0, 1/4)$ e cuja diretriz é a reta horizontal $y = -1/4$. Com efeito, a distância de um ponto qualquer (x, x^2) do gráfico de $f(x) = x^2$ ao ponto $F = (0, 1/4)$ é igual a

$$\sqrt{x^2 + (x^2 - 1/4)^2}.$$

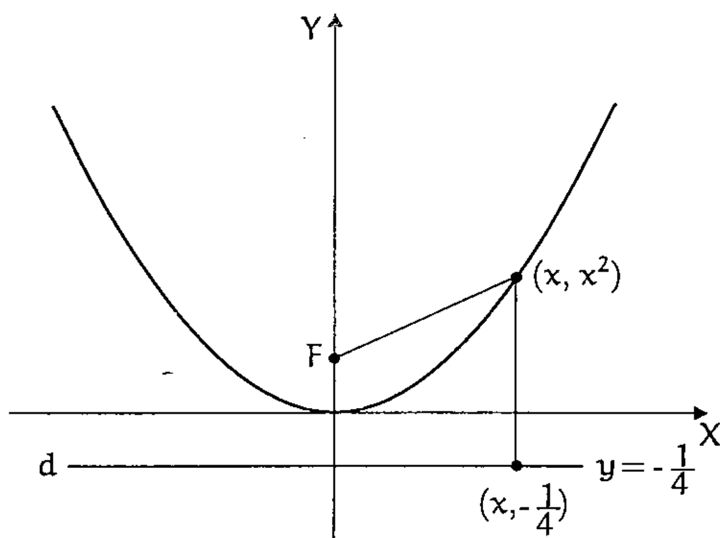


Figura 27

A distância do mesmo ponto (x, x^2) à reta $y = -1/4$ é $x^2 + 1/4$.

Como se trata de números positivos, para verificarmos a igualdade entre estas duas distâncias, basta ver que seus quadrados são iguais. E, de fato, tem-se

$$x^2 + (x^2 - 1/4)^2 = (x^2 + 1/4)^2,$$

para todo $x \in \mathbb{R}$, como se verifica facilmente. (Veja o Exercício 40.)

Exemplo 4. Se $a \neq 0$, o gráfico da função quadrática $f(x) = ax^2$ é a parábola cujo foco é $F = (0, 1/4a)$ e cuja diretriz é a reta horizontal $y = -1/4a$.

A fim de se convencer deste fato, basta verificar que, para todo $x \in \mathbb{R}$, vale a igualdade

$$x^2 + \left(ax^2 - \frac{1}{4a}\right)^2 = \left(ax^2 + \frac{1}{4a}\right)^2,$$

onde o primeiro membro é o quadrado da distância do ponto genérico $P = (x, ax^2)$ do gráfico de $f(x) = ax^2$ ao foco $F = (0, 1/4a)$ e o segundo membro é o quadrado da distância do mesmo ponto P à reta $y = -1/4a$.

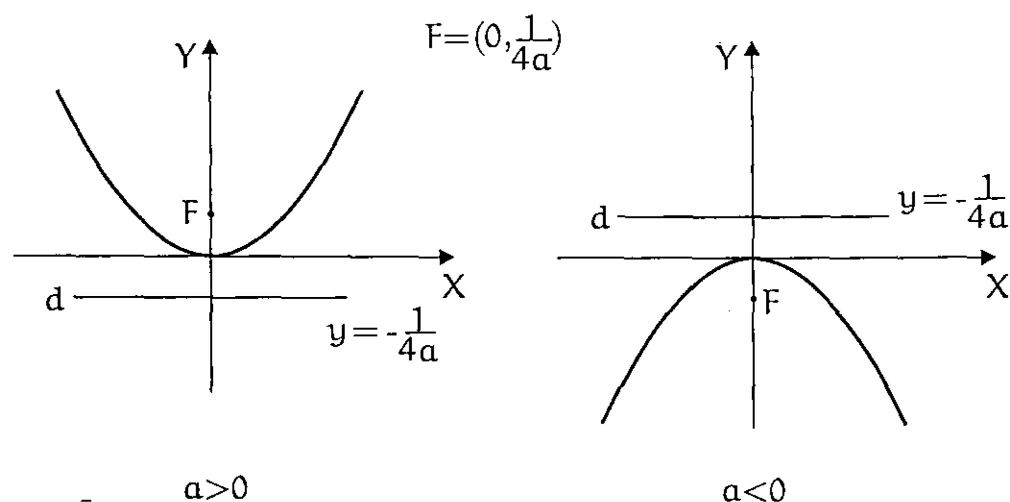


Figura 28

Conforme seja $a > 0$ ou $a < 0$, a parábola $y = ax^2$ tem sua concavidade voltada para cima ou para baixo.

Exemplo 5. Para todo $a \neq 0$ e todo $m \in \mathbb{R}$, o gráfico da função quadrática $f(x) = a(x - m)^2$ é uma parábola cujo foco é o ponto $F = (m, 1/4a)$ e cuja diretriz é a reta horizontal $y = -1/4a$.

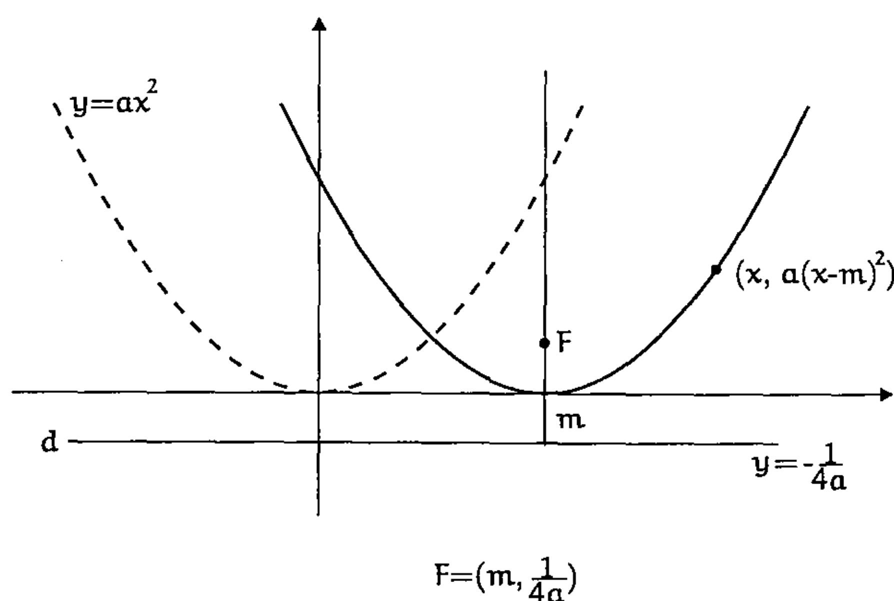


Figura 29

Para se chegar a esta conclusão, tem-se duas opções. Ou se verifica que, para todo $x \in \mathbb{R}$, vale a igualdade

$$(x-m)^2 + \left[a(x-m)^2 - \frac{1}{4a}\right]^2 = \left[a(x-m)^2 + \frac{1}{4a}\right]^2$$

ou então observa-se simplesmente que o gráfico de $f(x) = a(x-m)^2$ resulta do gráfico de $g(x) = ax^2$ pela translação horizontal $(x, y) \mapsto (x+m, y)$, a qual leva o eixo $x = 0$ no eixo $x = m$.

Exemplo 6. Dados $a, m, k \in \mathbb{R}$, com $a \neq 0$, o gráfico da função quadrática $f(x) = a(x-m)^2 + k$ é a parábola cujo foco é o ponto $F = \left(m, k + \frac{1}{4a}\right)$ e cuja diretriz é a reta horizontal $y = k - \frac{1}{4a}$.

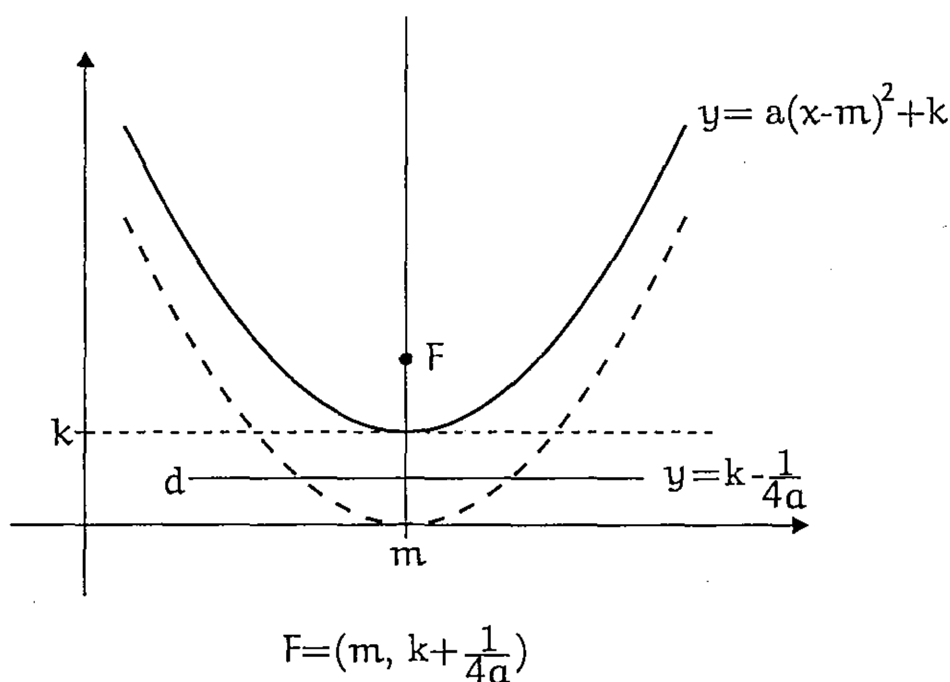


Figura 30

A afirmação acima resulta imediatamente do exemplo anterior, levando em conta que o gráfico da função quadrática $f(x) = a(x-m)^2 + k$ é obtido do gráfico de $g(x) = a(x-m)^2$ por meio da translação vertical $(x, y) \mapsto (x, y+k)$, que leva o eixo OX na reta $y = k$ e a reta $y = -1/4a$ na reta $y = k - \frac{1}{4a}$.

Segue-se deste último exemplo que o gráfico de qualquer

função quadrática

$$f(x) = ax^2 + bx + c$$

é uma parábola, cuja diretriz é a reta horizontal

$$y = \frac{4ac - b^2 - 1}{4a}$$

e cujo foco é o ponto

$$F = \left(-\frac{b}{2a}, \frac{4ac - b^2 + 1}{4a}\right).$$

Esta parábola tem sua concavidade voltada para cima se $a > 0$ ou para baixo se $a < 0$.

Com efeito, a forma canônica do trinômio

$$ax^2 + bx + c$$

nos dá

$$ax^2 + bx + c = a(x - m)^2 + k,$$

onde

$$m = -b/2a \quad \text{e} \quad k = (4ac - b^2)/4a.$$

O ponto do gráfico de

$$f(x) = ax^2 + bx + c$$

mais próximo da diretriz é aquele de abcissa $x = -b/2a$. Neste ponto, $f(x)$ atinge seu valor mínimo quando $a > 0$ e seu valor máximo quando $a < 0$. Ainda quando $x = -b/2a$, o ponto $(x, f(x))$ é o vértice da parábola que constitui o gráfico de $f(x)$.

A propriedade, provada no final da seção anterior, segundo a qual a função quadrática

$$f(x) = ax^2 + bx + c$$

assume valores iguais $f(x) = f(x')$ se, e somente se, os pontos x e x' são simétricos em relação a $-b/2a$ (ou seja, $x + x' = -b/a$) significa que a reta vertical $x = -b/2a$ é um eixo de simetria do gráfico de f; mais precisamente, é o eixo dessa parábola.

O gráfico da função quadrática

$$f(x) = ax^2 + bx + c$$

é um elemento de grande importância para entender o comportamento desta função. As abcissas α, β dos pontos onde esse gráfico intersecta o eixo OX são as raízes da equação

$$ax^2 + bx + c = 0.$$

O ponto médio do segmento $[\alpha, \beta]$ é a abcissa do vértice da parábola. Se o gráfico está inteiramente acima, ou inteiramente abaixo do eixo horizontal OX, a equação não possui raízes. Se o gráfico apenas tangencia o eixo OX, a equação tem uma raiz (única) dupla. Se $\alpha < x < \beta$ então $f(x)$ tem sinal contrário ao sinal de a; se $x < \alpha$ ou $x > \beta$, $f(x)$ tem o mesmo sinal de a. Estas e outras conclusões resultam imediatamente do exame do gráfico.

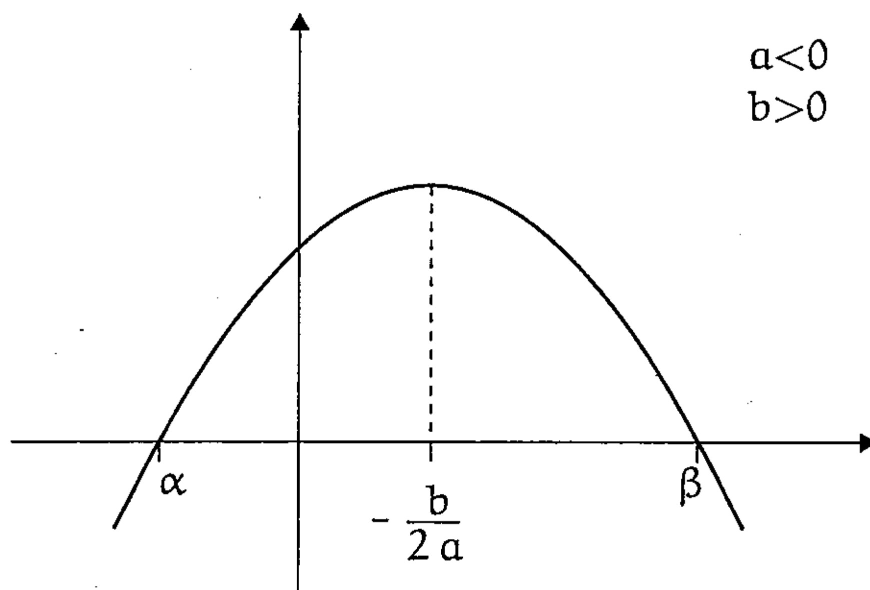


Figura 31

Examinaremos a seguir a questão de saber em que condições os gráficos de duas funções quadráticas, são parábolas congruentes. Começaremos com duas observações sobre gráficos, em geral.

1. *Aplicando a translação horizontal* $(x, y) \mapsto (x+m, y)$ *ao gráfico da função* $f: \mathbb{R} \to \mathbb{R}$, *obtém-se o gráfico da função* $g: \mathbb{R} \to \mathbb{R}$, *tal que* $g(x) = f(x - m)$ *para todo* $x \in \mathbb{R}$.

Com efeito, um ponto qualquer $(x, f(x))$ do gráfico de f é transformado por essa translação no ponto $(x + m, f(x))$. Escrevendo $\overline{x} = x + m$, donde $x = \overline{x} - m$, vemos que a translação considerada transforma cada ponto $(x, f(x))$ do gráfico f no ponto $(\overline{x}, f(\overline{x}-m)) = (\overline{x}, g(\overline{x}))$ do gráfico de g.

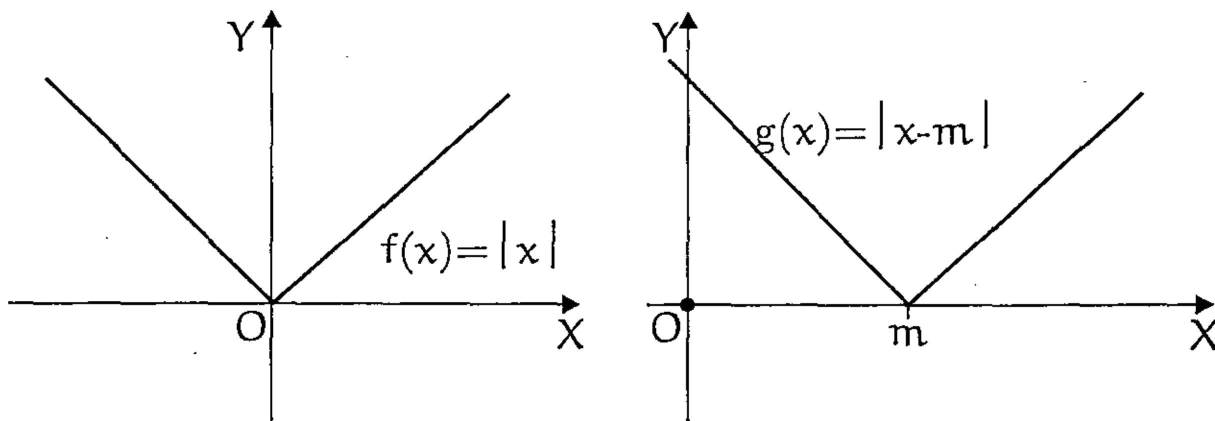


Figura 32

2. *A translação vertical* $(x, y) \mapsto (x, y + k)$ *transforma o gráfico da função* $f\colon \mathbb{R} \to \mathbb{R}$ *no gráfico da função* $h\colon \mathbb{R} \to \mathbb{R}$*, tal que* $h(x) = f(x) + k$ *para todo* $x \in \mathbb{R}$.

Com efeito, essa translação leva cada ponto $(x, f(x))$ do gráfico de f no ponto $(x, f(x) + k) = (x, g(x))$ do gráfico de g.

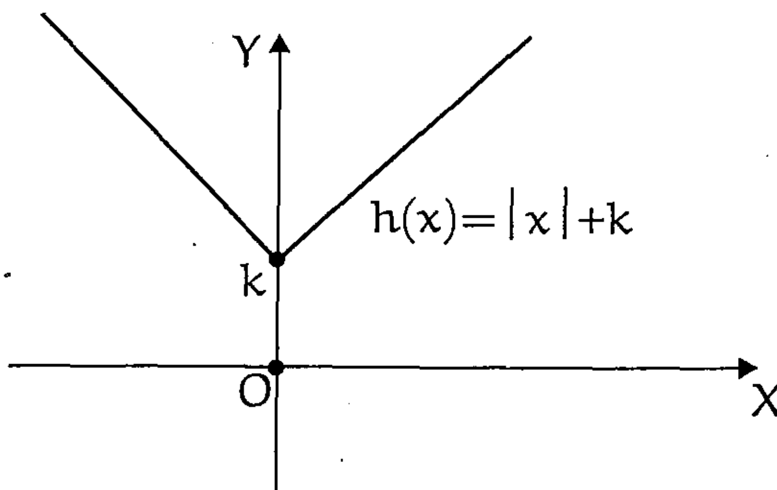


Figura 33

Consideremos agora, em particular, a função quadrática

$$f(x) = ax^2 + bx + c.$$

Sabemos que seu gráfico é uma parábola, cujo vértice tem abcissa igual a $m = -b/2a$. Submetendo essa parábola à translação horizontal $(x, y) \mapsto (x-m, y)$, obtemos uma nova parábola, cujo vértice tem abcissa igual a zero, isto é, está sobre o eixo OY. Pelo que vimos acima, esta nova parábola é o gráfico da função quadrática

$$\begin{aligned} g(x) = f(x-m) &= f\left(x - \frac{b}{2a}\right) \\ &= a\left(x - \frac{b}{2a}\right)^2 + b\left(x - \frac{b}{2a}\right) + c \\ &= ax^2 + k, \end{aligned}$$

onde

$$k = \frac{4ac - b^2}{4a}.$$

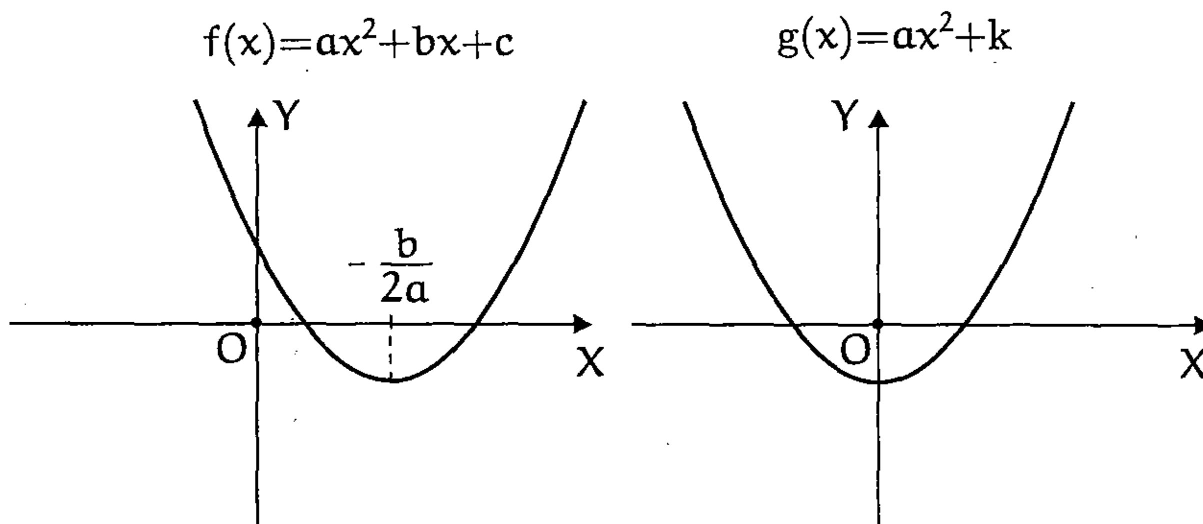


Figura 34

Em seguida, aplicamos a esta segunda parábola a translação vertical $(x, y) \mapsto (x, y-k)$, obtendo uma nova parábola, cujo vértice coincide com a origem $0 = (0, 0)$. Pela segunda obervação acima,

esta última parábola é o gráfico da função

$$h(x) = g(x) - k = ax^2 + k - k.$$

ou seja, $h(x) = ax^2$.

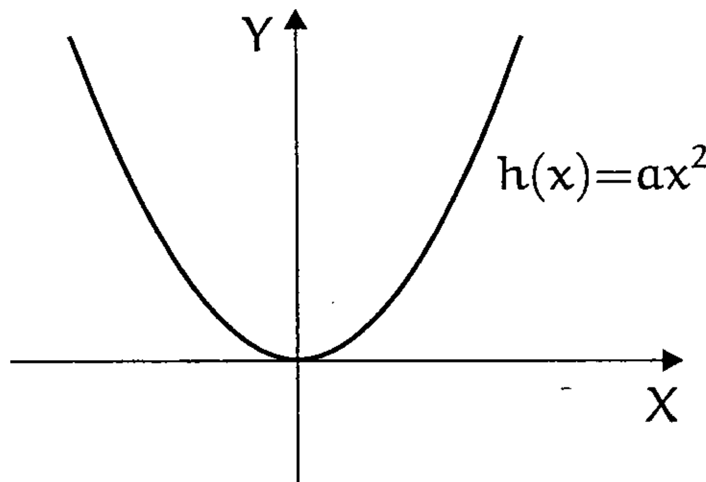


Figura 35

Pelo que acabamos de ver, a parábola que é o gráfico da função

$$f(x) = ax^2 + bx + c$$

transforma-se na parábola gráfico da função $h(x) = ax^2$ mediante uma translação horizontal seguida de uma translação vertical. Isto significa que essas duas parábolas são congruentes.

Assim, o gráfico da função

$$\varphi(x) = -ax^2 + bx + c$$

é congruente ao gráfico de $\psi(x) = -ax^2$. Por sua vez, a reflexão em torno do eixo horizontal, ou seja, a transformação $(x, y) \mapsto (x, -y)$, leva o gráfico de $\psi(x) = -ax^2$ no gráfico de $h(x) = ax^2$.

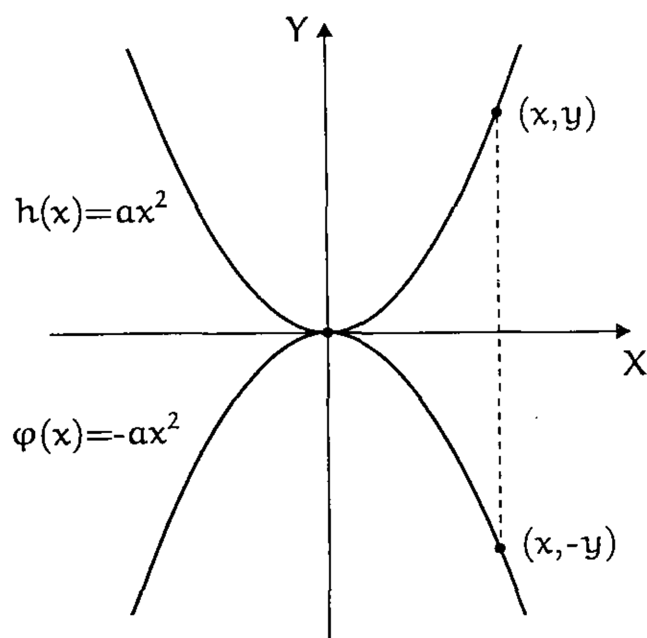


Figura 36

Podemos resumir a discussão acima enunciando: *Se* $a' = \pm a$ *então os gráficos das funções quadráticas* $f(x) = ax^2 + bx + c$ *e* $\varphi(x) = a'x^2 + b'x + c'$ *são parábolas congruentes.*

Quando $a' = a$, transformamos uma dessas parábolas na outra por meio de uma translação horizontal seguida de uma translação vertical. Se $a' = -a$, deve-se acrescentar ainda a reflexão em torno do eixo OX.

Vemos assim que, para a congruência das parábolas, gráficos das funções quadráticas $f(x) = ax^2+bx+c$ e $\varphi(x) = a'x^2+b'x+c'$, os coeficientes b, b' e c, c' não importam. Eles apenas determinam a posição da parábola em relação aos eixos: c é a ordenada do ponto em que a parábola corta o eixo vertical, enquanto b é a inclinação da tangente nesse mesmo ponto. (Vide pág. 137, logo adiante.)

Cabe, naturalmente, perguntar se os gráficos das funções f e φ podem ser congruentes, mesmo quando $a' \neq \pm a$. A resposta é negativa. Mais explicitamente, vale a recíproca do enunciado acima: se os gráficos das funções quadráticas $f(x) = ax^2 + bx + c$ e $\varphi(x) = a'x^2 + b'x + c'$ são parábolas congruentes então $a' = \pm a$.

Para mostrar isto, pelo que vimos acima, basta considerar as funções $f(x) = ax^2$ e $\varphi(x) = a'x^2$, com $a > 0$ e $a' > 0$. Se for $a < a'$

então $ax^2 < a'x^2$ (e se $a > a'$ então $ax^2 > a'x^2$) para todo $x \in \mathbb{R}$.

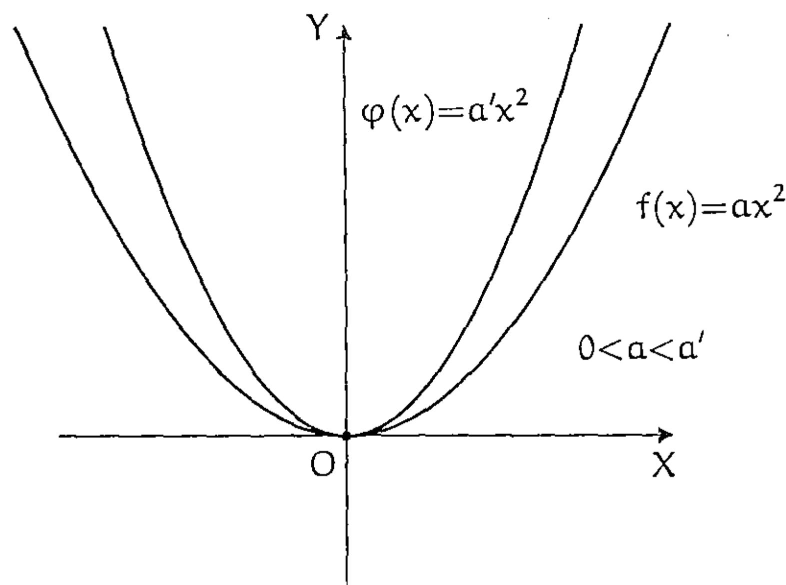


Figura 37

A figura acima deixa claro que as duas parábolas consideradas não são congruentes. Com efeito, duas parábolas com o mesmo vértice e o mesmo (semi-) eixo são como dois ângulos que têm o mesmo vértice e a mesma (semi-reta) bissetriz: só são congruentes se forem iguais, isto é, se coincidirem.

5. Uma Propriedade Notável da Parábola

Se girarmos uma parábola em torno do seu eixo, ela vai gerar uma superfície chamada *parabolóide de revolução*, também conhecida como *superfície parabólica*. Esta superfície possui inúmeras aplicações interessantes, todas elas decorrentes de uma propriedade geométrica da parábola, que veremos nesta seção.

A fama das superfícies parabólicas remonta à Antiguidade. Há uma lenda segundo a qual o extraordinário matemático grego Arquimedes, que viveu em Siracusa em torno do ano 250 A.C., destruiu a frota que sitiava aquela cidade incendiando os navios com os raios de sol refletidos em espelhos parabólicos. Embora isto seja teoricamente possível, há sérias dúvidas históricas sobre a capacidade tecnológica da época para fabricar tais espelhos.

Mas a lenda sobreviveu, e com ela a idéia de que ondas (de luz, de calor, de rádio ou de outra qualquer natureza), quando refletidas numa superfície parabólica, concentram-se sobre o foco, assim reforçando grandemente o sinal recebido.

Da lenda de Arquimedes restam hoje um interessante acendedor solar de cigarros e outros artefatos que provocam ignição fazendo convergir os raios de sol para o foco de uma superfície parabólica polida.

Outros instrumentos atuam inversamente, concentrando na direção paralela ao eixo os raios de luz que emanam do foco. Como exemplos, citamos os holofotes, os faróis de automóveis e as simples lanternas de mão, que têm fontes luminosas à frente de uma superfície parabólica refletora.

Um importante uso recente destas superfícies é dado pelas antenas parabólicas, empregadas na rádio-astronomia, bem como no dia-a-dia dos aparelhos de televisão, refletindo os débeis sinais provenientes de um satélite sobre sua superfície, fazendo-os convergir para um único ponto, o foco, deste modo reforçando-os consideravelmente.

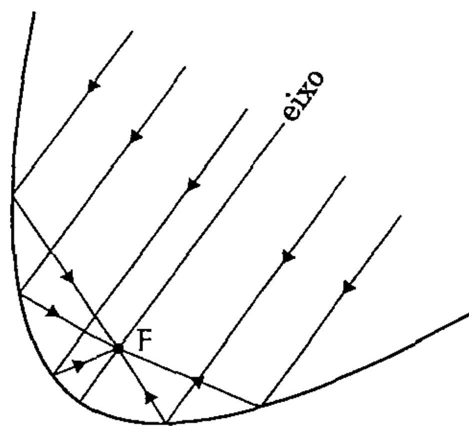


Figura 38

Vamos agora analisar o fundamento matemático desses aparelhos.

Começaremos com o princípio segundo o qual, quando um raio incide sobre uma superfície refletora, o ângulo de incidência é igual ao ângulo de reflexão.

Neste contexto, a superfície parabólica pode ser substituída pela parábola que é a interseção dessa superfície com o plano que contém o raio incidente, o raio refletido e o eixo de rotação (igual ao eixo da parábola).

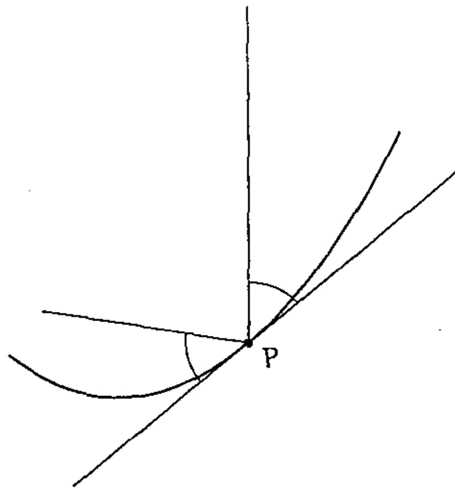


Figura 39

O ângulo entre uma reta e uma curva que se intersetam no ponto P é, por definição, o ângulo entre essa reta e a tangente à curva traçada pelo ponto de interseção. É assim que se interpretam os ângulos de incidência e reflexão.

A *tangente* a uma parábola no ponto P é a reta que tem em comum com a parábola esse único ponto P e tal que todos os demais pontos da parábola estão do mesmo lado dessa reta.

A tangente a uma parábola tem sua posição determinada pelo teorema seguinte.

Se a parábola é o gráfico da função $f(x) = ax^2 + bx + c$, *sua tangente no ponto* $P = (x_0, y_0)$, *onde* $y_0 = ax_0^2 + bx_0 + c$, *é a reta que passa por esse ponto e tem inclinação igual a* $2ax_0 + b$.

Para provar isto, mostremos que todos os pontos dessa parábola que têm abcissa diferente de x_0 estão fora da reta mencionada e no mesmo semi-plano determinado por ela.

Mais precisamente, suponhamos (para fixar idéias) que seja $a > 0$. Mostraremos que, para todo $x \neq x_0$, o ponto (x, y) da parábola, com $y = ax^2 + bx + c$ está acima do ponto $(x, y_0 + (2ax_0 +$

$b)(x - x_0))$, de mesma abcissa x, situado sobre a reta. Noutras palavras, queremos provar que (supondo $a > 0$)

$$x \neq x_0 \Rightarrow ax^2 + bx + c > ax_0^2 + bx_0 + c + (2ax_0 + b)(x - x_0).$$

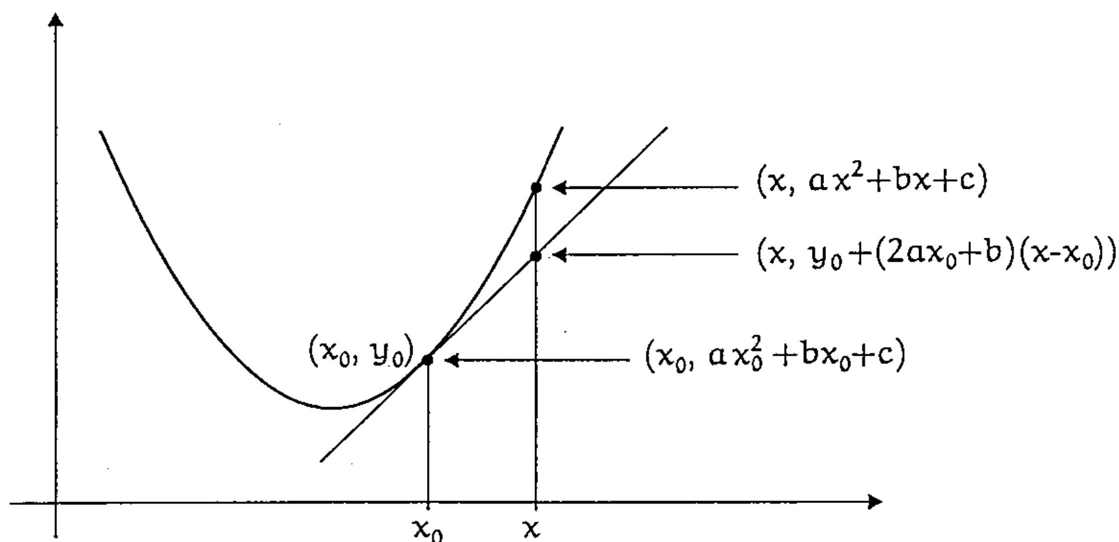


Figura 40

Para isto, basta notar que

$$x \neq x_0$$
$$\Downarrow$$
$$ax^2 + bx + c - [ax_0^2 + bx_0 + c + (2ax_0 + b)(x - x_0)] =$$
$$= a(x - x_0)^2 > 0.$$

Isto mostra que a reta de inclinação $2ax_0 + b$ que passa pelo ponto (x_0, y_0), com $y_0 = f(x_0)$, tem este único ponto em comum com a parábola que é o gráfico de f e que todos os pontos da parábola estão acima dessa reta. Logo esta reta é tangente à parábola neste ponto. Quando $a > 0$, a parábola se situa acima de qualquer de suas tangentes, conforme acabamos de ver. Se for $a < 0$ então a parábola se situa abaixo de todas as suas tangentes.

Observação: Todas as retas paralelas ao eixo de uma parábola têm apenas um ponto em comum com essa parábola mas nenhuma

delas é tangente porque há pontos da parábola em ambos semiplanos por ela determinados.

Sabendo que a parábola, gráfico da função

$$f(x) = ax^2 + bx + c,$$

tem, no ponto $P = (x, y)$, uma tangente cuja inclinação é $2ax + b$, calculemos agora a inclinação da reta FQ que une o foco F ao ponto Q, pé da perpendicular baixada de P sobre a diretriz d.

No cálculo que se segue, admitiremos que P não é o vértice da parábola, isto é, que sua abcissa x é diferente de $-b/2a$, logo $2ax + b \neq 0$. Caso P fosse o vértice, a reta FQ seria vertical e a tangente no ponto P teria inclinação zero, logo seria horizontal.

A inclinação da reta FQ é dada por uma fração cujo numerador é a diferença entre as ordenadas de Q e F e cujo denominador é a diferença entre as abcissas desses pontos.

Ora, já vimos que $F = \left(m, k + \frac{1}{4a}\right)$ e $Q = \left(x, k - \frac{1}{4a}\right)$, onde $m = -b/2a$ e k = ordenada do vértice da parábola. Logo a inclinação de FQ é igual a

$$\frac{k - \frac{1}{4a} - \left(k + \frac{1}{4a}\right)}{x - m} = \frac{-1}{2a(x - m)} = \frac{-1}{2a\left(x + \frac{b}{2a}\right)} = -\frac{1}{2ax + b}.$$

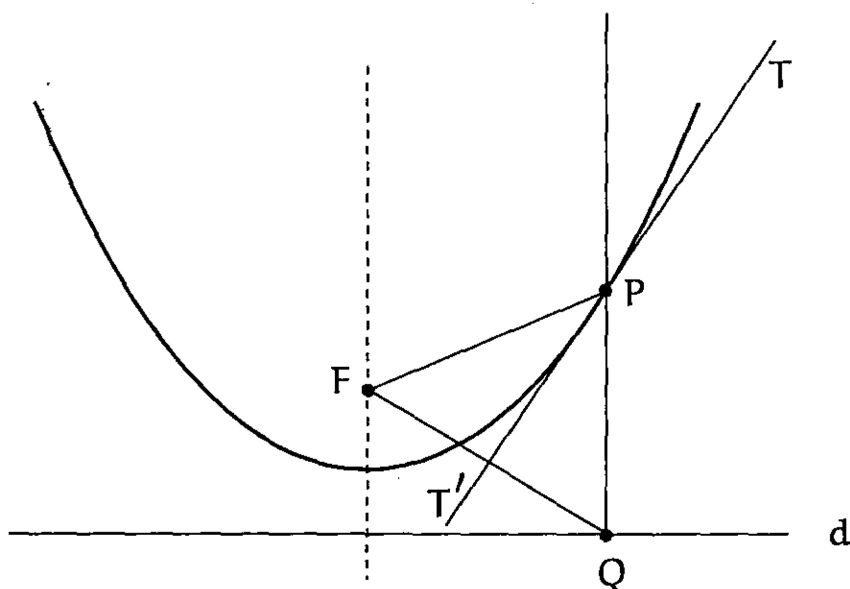


Figura 41

Isto significa que o segmento de reta FQ é perpendicular à reta TT', tangente à parábola no ponto P, conforme o

Lema: *As retas* $y = ax + b$ *e* $y = a'x + b'$, *com* $a \neq 0$ *e* $a' \neq 0$, *são perpendiculares se, e somente se,* $a' = -1/a$.

Demonstração: Como as retas $y = ax$ e $y = a'x$ são paralelas às retas dadas, aquelas serão perpendiculares se, e somente se, estas o forem. Suponhamos que estas retas sejam perpendiculares. Tomando $x = 1$, vemos que o ponto $(1, a)$ pertence a uma das retas e o ponto $(1, a')$ pertence à outra (ver figura na próxima página).

Então o triângulo cujos vértices são os pontos $(0, 0)$, $(1, a)$ e $(1, a')$ é retângulo, logo a altura baixada do vértice do ângulo reto é a média geométrica dos segmentos que ela determina sobre a hipotenusa. Ora, o comprimento da altura é 1. Além disso, um dos números a e a' (digamos a') é negativo e o outro é positivo. Logo os referidos segmentos medem a e $-a'$. Assim $1 = -aa'$ e $a' = -1/a$. Reciprocamente, se $a' = -1/a$, consideramos a reta $y = bx$, perpendicular à reta $y = ax$ a partir da origem. Pelo que acabamos de ver, temos $b = -1/a$, logo $b = a'$, logo $y = a'x$ coincide com $y = bx$ portanto é perpendicular a $y = ax$.

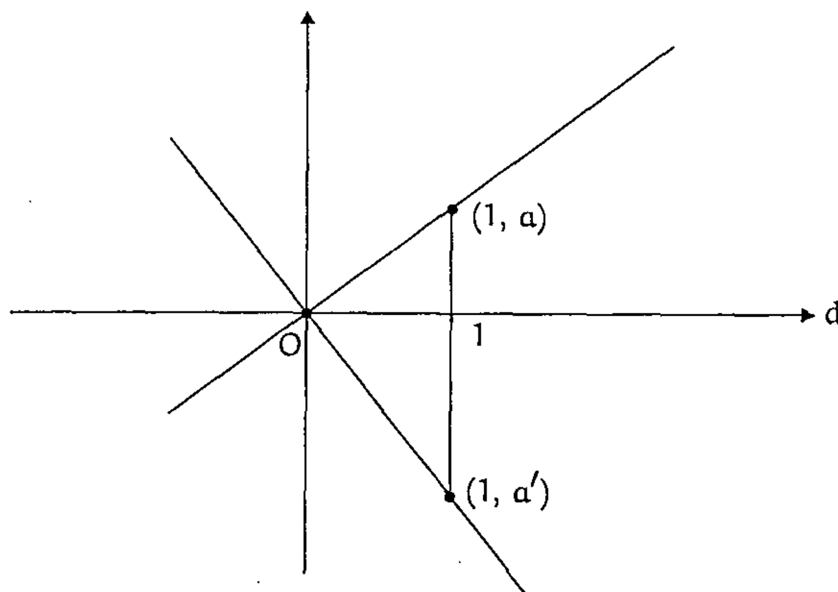


Figura 42

Podemos, finalmente, enunciar a propriedade geométrica da

parábola na qual se baseiam as aplicações da superfície parabólica.

A tangente à parábola num ponto P *faz ângulos iguais com a paralela ao eixo e com a reta que une o foco* F *a esse ponto.*

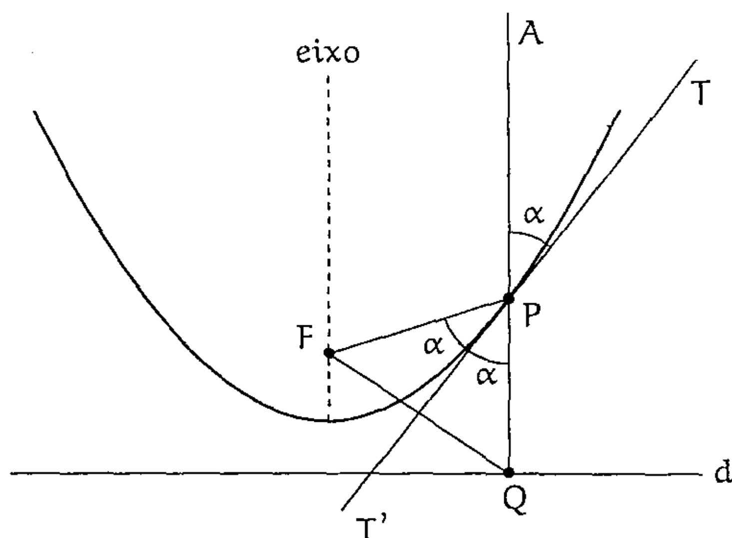


Figura 43

Com efeito, se Q é o pé da perpendicular baixada de P sobre a diretriz, a definição da parábola nos diz que $\overline{FP} = \overline{PQ}$, logo o triângulo FPQ é isósceles. Além disso, acabamos de ver que FQ é perpendicular à tangente, ou seja, a tangente é altura desse triângulo isósceles, logo é também bissetriz. Portanto, os ângulos $F\widehat{P}T'$ e $T'\widehat{P}Q$ são iguais. Logo $A\widehat{P}T = F\widehat{P}T' = \alpha$.

Se a antena parabólica estiver voltada para a posição (estacionária) do satélite, a grande distância faz com que os sinais emitidos por este sigam trajetórias praticamente paralelas ao eixo da superfície da antena, logo eles se refletem na superfície e convergem para o foco, de acordo com o princípio que acabamos de demonstrar.

6. O Movimento Uniformemente Variado

A função quadrática é o modelo matemático que descreve o movimento uniformemente variado.

Neste tipo de movimento, que tem como um exemplo importante a queda dos corpos no vácuo, sujeitos apenas à ação da

gravidade, tem-se um ponto que se desloca sobre um eixo. Sua posição no instante t é dada pela abcissa f(t). O que caracteriza o movimento uniformemente variado é o fato de f ser uma função quadrática:

$$(*) \qquad f(t) = \frac{1}{2}at^2 + bt + c.$$

Nesta expressão a constante a chama-se a *aceleração*, b é a *velocidade inicial* (no instante $t = 0$) e c é a *posição inicial* do ponto.

Em qualquer movimento, dado por uma função f, o quociente

$$\frac{f(t+h) - f(t)}{h} = \frac{\text{espaço percorrido}}{\text{tempo de percurso}}$$

chama-se a *velocidade média* do ponto no intervalo cujos extremos são t e $t+h$. No caso em que f é dada pela fórmula (*), a velocidade média do móvel entre os instantes t e $t+h$ é igual a $at + b + \frac{ah}{2}$. Se tomarmos h cada vez menor, este valor se aproxima de $at + b$. Por isso se diz que

$$v(t) = at + b$$

é a *velocidade* do ponto (no movimento uniformemente variado) no instante t.

Quando $t = 0$ temos $v(0) = b$, por isso b se chama a velocidade inicial. Além disso, vê-se que $a = [v(t+h) - v(t)]/h$ para quaisquer t, h, logo a aceleração constante a é a taxa de variação da velocidade. Por isso o movimento se chama uniformemente variado. [Uniformemente acelerado ou retardado, conforme v tenha o mesmo sinal de a (isto é, $t > -b/a$) ou v tenha sinal oposto ao de a (ou seja, $t < -b/a$).]

No caso da queda livre de um corpo, a aceleração a é a da gravidade, normalmente indicada pela letra g.

Nosso conhecimento da função quadrática permite obter uma descrição completa do movimento uniformemente variado.

Por exemplo, se uma partícula é posta em movimento sobre um eixo a partir do ponto de abcissa -6, com velocidade inicial de $5m/seg$ e aceleração constante de $-2m/seg^2$, quanto tempo se

passa até sua trajetória mude de sentido e ela comece a voltar para o ponto de partida? Resposta: temos $f(t) = -t^2 + 5t - 6$. Logo o valor máximo de f é obtido quando $t = -5/(-2) = 2{,}5seg$. Podemos ainda dizer que o ponto começa a voltar quando $v(t) = 0$. Como $v(t) = -2t + 5$ isto nos dá novamente $t = 2{,}5seg$.

O movimento uniformemente variado pode ocorrer também no plano. Um exemplo disso é o movimento de um projétil (uma bala, uma bola, uma pedra, etc.) lançado por uma força instantânea e, a partir daí, sujeito apenas à ação da gravidade, sendo desprezada a resistência do ar (movimento no vácuo). Embora o processo ocorra no espaço tridimensional, a trajetória do projétil está contida no plano determinado pela reta vertical no ponto de partida e pela direção da velocidade inicial.

Quando se tem um movimento retilíneo (sobre um eixo), a velocidade do móvel é expressa por um número. Mas quando o movimento ocorre no plano ou no espaço, a velocidade é expressa por um vetor (segmento de reta orientado), cujo comprimento se chama a *velocidade escalar* do móvel (tantos metros por segundo). A direção e o sentido desse vetor indicam a direção e o sentido do movimento.

No plano em que se dá o movimento, tomemos um sistema de coordenadas cuja origem é o ponto de partida do projétil e cujo eixo OY é a vertical que passa por esse ponto.

A velocidade inicial do projétil é o vetor $v = (v_1, v_2)$ cuja primeira coordenada v_1 fornece a velocidade da componente horizontal do movimento (deslocamento da sombra, ou projeção do projétil sobre o eixo horizontal OX).

Como a única força atuando sobre o projétil é a gravidade, a qual não possui componente horizontal, nenhuma força atua sobre este movimento horizontal, que é portanto um movimento uniforme. Assim, se $P = (x, y)$ é a posição do projétil no instante t, tem-se $x = v_1 t$.

Por sua vez, a aceleração (= força) da gravidade é constante, vertical, igual a $-g$. (O sinal menos se deve ao sentido da gra-

vidade ser oposto à orientação do eixo vertical OY.) Portanto, a componente vertical do movimento de P é um movimento uniformemente acelerado sobre o eixo OY, com aceleração igual a $-g$ e velocidade inicial v_2.

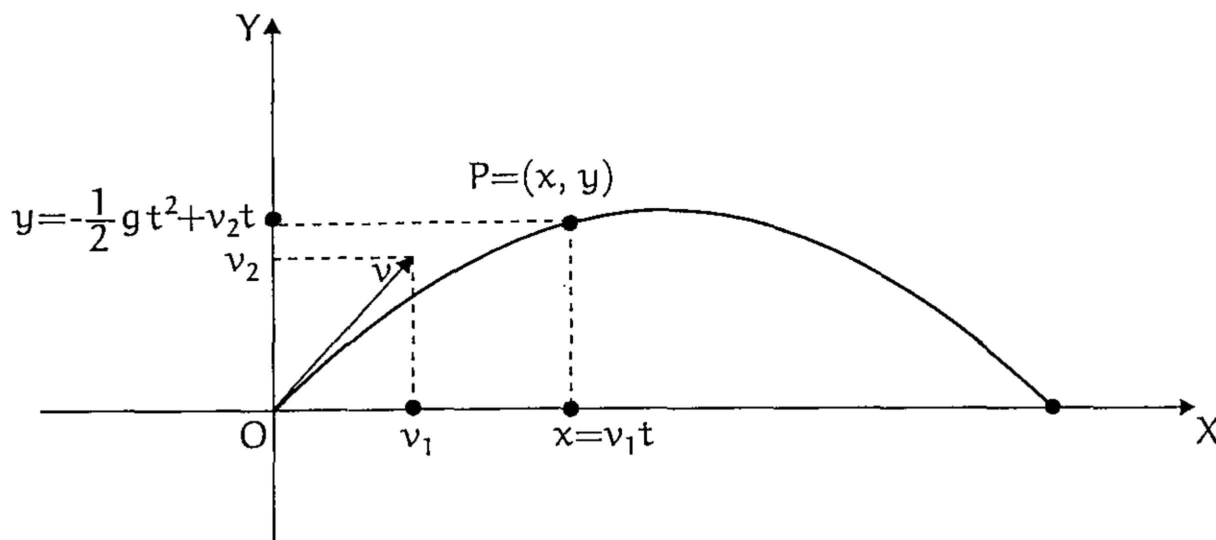


Figura 44

Logo, em cada instante t, a ordenada y do ponto $P = (x, y)$ é dada por $y = -\frac{1}{2}gt^2 + v_2 t$. (Não há termo constante porque $y = 0$ quando $t = 0$.)

Se $v_1 = 0$ então, para todo t, tem-se $x = v_1 t = 0$, logo $P = (0, y)$, com

$$y = -\frac{1}{2}gt^2 + v_2 t.$$

Neste caso, a trajetória do projétil é vertical.

Suponhamos agora $v_1 \neq 0$. Então, de $x = v_1 t$ vem $t = x/v_1$. Substituindo t por este valor na expressão de y, obtemos

$$y = ax^2 + bx, \quad \text{onde} \quad a = -g/2v_1^2 \quad \text{e} \quad b = v_2/v_1.$$

Isto mostra que a trajetória do projétil é uma parábola.

7. Caracterização das Funções Quadráticas

A função quadrática mais simples, $f(x) = x^2$, transforma a progressão aritmética

$$1, 2, 3, 4, \ldots, n, n+1, \ldots$$

na seqüência

$$1, 4, 9, 16, \ldots, n^2, n^2 + 2n + 1, \ldots,$$

que não é uma progressão aritmética, ou seja, a diferença entre dois termos consecutivos não é constante. Entretanto, se examinarmos as diferenças entre os termos consecutivos desta última seqüência, encontraremos

$$3, 5, 7, \ldots, 2n + 1, \ldots,$$

que é uma progressão aritmética.

Isto não é uma coincidência. Se $f(x) = ax^2 + bx + c$ é uma função quadrática arbitrária e

$$x_1, x_2, x_3, x_4, \ldots$$

é uma progressão aritmética qualquer então a seqüência

$$y_1, y_2, y_3, y_4, \ldots$$

dos valores $y_1 = f(x_1)$, $y_2 = f(x_2)$, $f_3 = f(x_3)$, etc. goza da propriedade de que as diferenças sucessivas

$$d_1 = y_2 - y_1, \quad d_2 = y_3 - y_2, \quad d_3 = y_4 - y_3, \ldots$$

formam uma progressão aritmética. Mais precisamente, se $x_{i+1} - x_i = r$ para todo $i = 1, 2, 3, \ldots$ então $d_{i+1} - d_i = 2ar^2$. Este fato, que se verifica sem maiores dificuldades, constitui uma propriedade exclusiva das funções quadráticas, como veremos a seguir.

No caso do movimento uniformemente acelerado, se considerarmos a queda livre de um corpo, sujeito apenas à ação da gravidade, pode-se verificar experimentalmente que, marcando a posição do corpo em intervalos iguais e sucessivos de tempo (digamos, de segundo em segundo), a partir do início da queda, as distâncias percorridas em cada intervalo de um segundo vão crescendo, e formam uma progressão aritmética de razão g, onde $g = 9,8\text{m}/\text{seg}^2$ é a aceleração da gravidade (ver Figura 45). O teorema de caracterização que provaremos logo a seguir garante

então que a altura $f(t)$ do corpo em queda livre depois de t segundos do início da queda é uma função quadrática: $f(t) = A - \frac{1}{2}gt^2$, onde A é a altura do ponto onde teve início a queda.

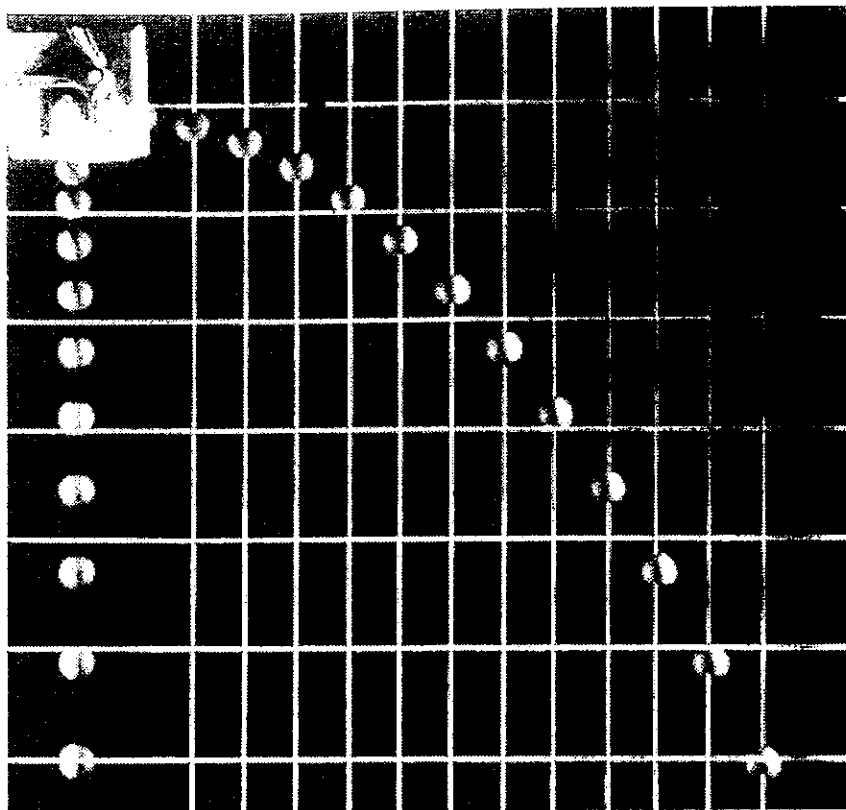


Figura 45

Foi estabelecido no capítulo anterior, como conseqüência do Teorema Fundamental da Proporcionalidade, que uma função crescente (ou decrescente) $\varphi: \mathbb{R} \to \mathbb{R}$ é afim se, e somente se, transforma progressões aritméticas em progressões aritméticas. Foi também observado, no final da Seção 3 daquele capítulo, que a monotonicidade da função φ pode ser substituída pela continuidade dessa mesma função.

Como uma função quadrática nunca pode ser monótona, nos teoremas de caracterização que apresentamos a seguir, trabalharemos com a hipótese de continuidade em vez de monotonicidade. Admitiremos conhecido que uma função quadrática é contínua e que se duas funções contínuas $f, g: \mathbb{R} \to \mathbb{R}$ são tais que $f(r) = g(r)$ para todo racional r então $f(x) = g(x)$ para todo x real.

Uma *progressão artitmética de segunda ordem* é uma seqüência $y_1, y_2, y_3, y_4, \ldots$ tal que as diferenças sucessivas

$$d_1 = y_2 - y_1, \quad d_2 = y_3 - y_2, \quad d_3 = y_4 - y_3, \ldots$$

formam uma progressão aritmética usual.

Por exemplo, a seqüência $1, 4, 9, 16, 25, \ldots$ dos quadrados dos números naturais é uma progressão aritmética de segunda ordem. Isto significa que a função quadrática $f(x) = x^2$ transforma a progressão aritmética $1, 2, 3, 4, 5, \ldots$ na progressão aritmética de segunda ordem $f(1), f(2), f(3), f(4), f(5), \ldots$. Mais geralmente, como vimos acima, se $f\colon \mathbb{R} \to \mathbb{R}$ é qualquer função quadrática e $x_1, x_2, x_3, x_4, \ldots$ é uma progressão aritmética arbitrária então os números $y_1 = f(x_1), y_2 = f(x_2), y_3 = f(x_3), y_4 = f(x_4), \ldots$ formam uma progressão aritmética de segunda ordem.

Mostraremos a seguir que, reciprocamente, toda função contínua $f\colon \mathbb{R} \to \mathbb{R}$ que transforma progressões aritméticas em progressões aritméticas de segunda ordem é da forma $f(x) = ax^2 + bx + c$.

Começaremos lembrando que uma progressão aritmética é a restrição de uma função afim aos números naturais: se $x_1, x_2, \ldots, x_n, \ldots$ é uma P.A. de razão r então a igualdade $x_n = x_1 + (n-1)r$ pode ser escrita como $x_n = an + b$, onde $a = r$ e $b = x_1 - r$. Logo a função afim $f(x) = ax + b$, quando restrita aos números naturais, fornece os termos $x_1 = f(1), x_2 = f(2), \ldots, x_n = f(n), \ldots$ da P.A.

De modo análogo, conforme veremos agora, se $y_1, y_2, \ldots, y_n, \ldots$ é uma P.A. de segunda ordem, existem números reais a, b, c tais que $y_n = an^2 + bn + c$ para todo $n \in \mathbb{N}$. Assim, considerando a função $f(x) = ax^2 + bx + c$, temos $y_n = f(n)$ para todo $n \in \mathbb{N}$, portanto a restrição de f aos números naturais fornece os termos da P.A. de segunda ordem dada.

Com efeito, as diferenças sucessivas

$$y_2 - y_1, y_3 - y_2, \ldots, y_{n+1} - y_n, \ldots$$

formam uma P.A. ordinária, cujo primeiro termo é $d = y_2 - y_1$ e

cuja razão chamaremos de r; portanto seu n-ésimo termo é

$$y_{n+1} - y_n = d + (n-1)r,$$

para $n = 1, 2, 3, \ldots$. Temos então:

$$\begin{aligned} y_{n+1} &= (y_{n+1} - y_n) + (y_n - y_{n-1}) + \cdots + (y_3 - y_2) + y_1; \\ &= [d + (n-1)r] + [d + (n-2)r] + \cdots + [d + r] + d + y_1; \\ &= nd + \frac{n(n-1)}{2} r + y_1, \end{aligned}$$

para todo $n \in \mathbb{N}$.

Esta igualdade é igualmente verdadeira quando $n = 0$, o que nos permite escrever

$$\begin{aligned} y_n &= (n-1)d + \frac{(n-1)(n-2)}{2} r + y_1; \\ &= \frac{r}{2} n^2 + (d - \frac{3r}{2})n + r - d + y_1; \\ &= an^2 + bn + c, \end{aligned}$$

para todo $n \in \mathbb{N}$, com $a = r/2$, $b = d - 3r/2$, $c = r - d + y_1$.

Exemplo: A seqüência $3, 7, 13, 21, 31, 43, \ldots$, é uma P.A. de segunda ordem, pois as diferenças sucessivas $7-3$, $13-7$, $21-13$, $31-21$, $43-31$, $\ldots$, formam a P.A. ordinária $4, 6, 8, 10, 12, \ldots$, de razão $r = 2$ e primeiro termo $d = 4$. Segue-se do que acabamos de ver que o n-ésimo termo da seqüência inicial é $y_n = an^2 + bn + c$, onde $a = r/2 = 1$, $b = d - 3r/2 = 4 - 3 = 1$ e $c = r - d + y_1 = 2 - 4 + 3 = 1$. Noutras palavras, o termo de ordem n da seqüência $3, 7, 13, 21, \ldots$ é $y_n = n^2 + n + 1$.

Observação: Uma P.A. pode ter razão $x_{n+1} - x_n = 0$. Neste caso, trata-se de uma seqüência constante: $x_1, x_1, x_1, \ldots$. Consequentemente, uma P.A. de segunda ordem pode reduzir-se a uma P.A. ordinária, quando a razão r da P.A. $y_2 - y_1, y_3 - y_2, \ldots$ for igual a zero. Neste caso, $a = r/2 = 0$ e a função $f(x) = ax^2 + bx + c$, com $y_n = f(n)$, não é quadrática, reduzindo-se a $f(x) = bx + c$.

No teorema de caracterização, que demonstraremos a seguir, a fim de obtermos uma função quadrática, precisamos supor que a P.A. de segunda ordem que ocorre em seu enunciado é *não-degenerada*, isto é, não é uma P.A. ordinária.

Teorema (Caracterização das Funções Quadráticas.) *A fim de que a função contínua* $f: \mathbb{R} \to \mathbb{R}$ *seja quadrática é necessário e suficiente que toda progressão aritmética não-constante* $x_1, x_2, \ldots, x_n, \ldots$ *seja transformada por* f *numa progressão aritmética de segunda ordem não-degenerada* $y_1 = f(x_1), y_2 = f(x_2), \ldots y_n = f(x_n), \ldots$.

Demonstração: A necessidade já foi demonstrada acima. Para provar a suficiência, seja $f: \mathbb{R} \to \mathbb{R}$ uma função contínua com a propriedade de transformar toda P.A. não-constante numa P.A. de segunda ordem não-degenerada. Substituindo $f(x)$ por $g(x) = f(x) - f(0)$, vemos que g tem as mesmas propriedades de f e mais a propriedade adicional de que $g(0) = 0$.

Considerando a progressão aritmética $1, 2, 3, 4, 5, \ldots$, vemos que os valores $g(1), g(2), \ldots, g(n) \ldots$ formam uma P.A. de segunda ordem não-degenerada. Logo existem constantes $a \neq 0$ e b tais que

$$g(n) = an^2 + bn$$

para todo $n \in \mathbb{N}$. (Deveria ser $g(n) = an^2 + bn + c$ porém $g(0) = 0$.)

Em seguida, fixemos arbitrariamente um número $p \in \mathbb{N}$ e consideremos a progressão aritmética

$$\frac{1}{p}, \frac{2}{p}, \frac{3}{p}, \ldots, \frac{n}{p}, \ldots.$$

De modo análogo, concluímos que existem $a' \neq 0$ e b' tais que

$$g(\frac{n}{p}) = a'n^2 + b'n$$

para todo $n \in \mathbb{N}$. Assim, para todo $n \in \mathbb{N}$, temos:

$$\begin{aligned} an^2 + bn &= g(n) \\ &= g(\frac{np}{p}) \\ &= a'(np)^2 + b'(np) \\ &= (a'p^2)n^2 + (b'p)n. \end{aligned}$$

Portanto as funções quadráticas

$$ax^2 + bx \quad \text{e} \quad (a'p^2)x^2 + (b'p)x$$

coincidem para todo $x = n \in \mathbb{N}$. Como vimos no início deste capítulo, isto obriga a $a = a'p^2$ e $b = b'p$, ou seja, $a' = a/p^2$, $b' = b/p$. Logo, para quaisquer números naturais n e p vale:

$$\begin{aligned} g(\frac{n}{p}) &= a'n^2 + b'n \\ &= \frac{a}{p^2}n^2 + \frac{b}{p}n \\ &= a\left(\frac{n}{p}\right)^2 + b\left(\frac{n}{p}\right). \end{aligned}$$

Vemos então que as funções contínuas $g(x)$ e $ax^2 + bx$ são tais que $g(r) = ar^2 + br$ para todo número racional positivo $r = n/p$. Segue-se que $g(x) = ax^2 + bx$ para todo número real positivo x. De modo análogo, considerando a P.A. $-1, -2, -3, \ldots$, concluiríamos que $g(x) = ax^2 + bx$ para todo $x \leqslant 0$. Logo, pondo $f(0) = c$, temos $f(x) = g(x) + c$, ou seja

$$f(x) = ax^2 + bx + c$$

para todo $x \in \mathbb{R}$.

Exercícios

1. Encontre a função quadrática cujo gráfico é dado em cada figura abaixo:

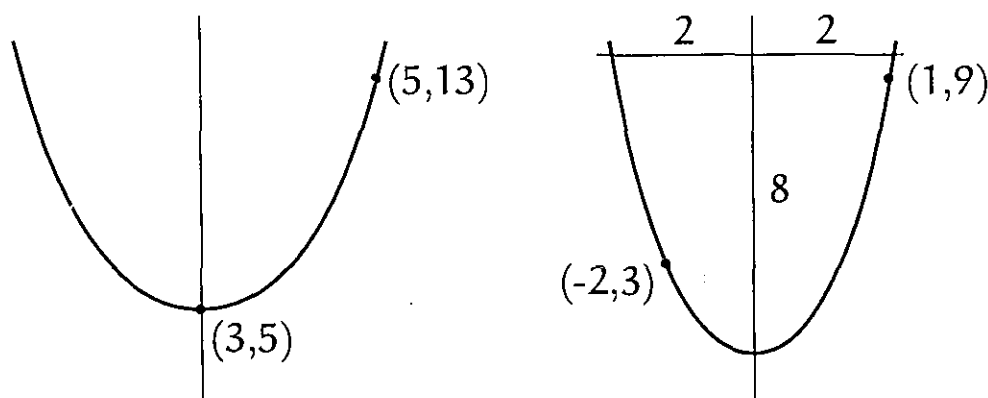


Figura 46

2. Identifique os sinais de a, b e c nos gráficos de funções quadráticas $f(x) = ax^2 + bx + c$ dados abaixo:

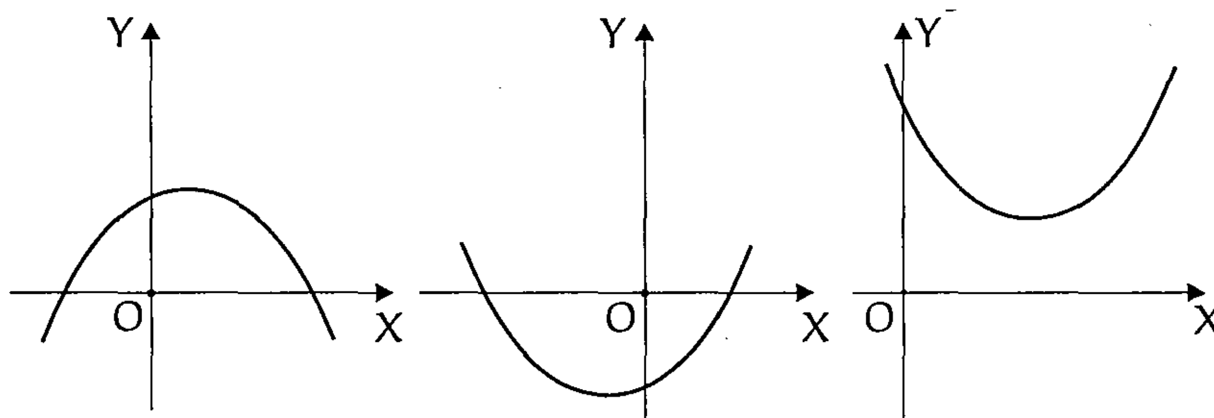


Figura 47

3. Escreva cada uma das funções quadráticas abaixo na forma $f(x) = a(x - m)^2 + k$. A seguir, calcule suas raízes (se existirem), o eixo de simetria de seu gráfico e seu valor mínimo ou máximo

a) $f(x) = x^2 - 8x + 23$

b) $f(x) = 8x - 2x^2$

4. Observe os gráficos abaixo, que representam as parábolas $y = ax^2$ para diversos valores de a. Estas parábolas são semelhantes entre si?

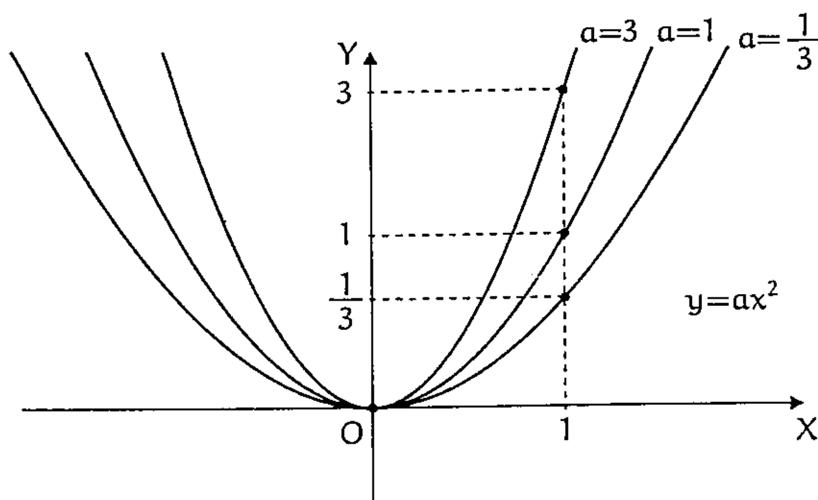


Figura 48

5. Encontre a unidade que deve ser usada nos eixos cartesianos de modo que a parábola abaixo seja o gráfico da função $f(x) = 2x^2$.

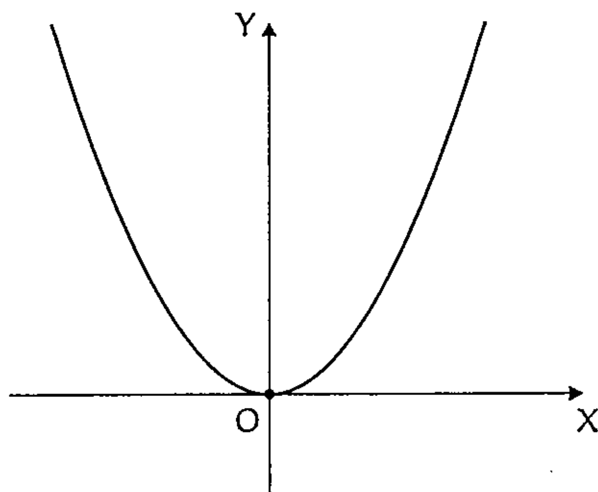


Figura 49

6. Encontre os valores mínimo e máximo assumidos pela função $f(x) = x^2 - 4x + 3$ em cada um dos intervalos abaixo:

a) $[1, 4]$

b) $[6, 10]$

7. Seja $f(x) = ax^2 + bx + c$, com $a > 0$.

a) Mostre que

$$f\left(\frac{x_1 + x_2}{2}\right) < \frac{f(x_1) + f(x_2)}{2}.$$

b) Mais geralmente, mostre que se $0 < \alpha < 1$, então

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha)f(x_2).$$

Interprete geometricamente esta propriedade.

8. Prove que se a, b e c são inteiros ímpares, as raízes de $y = ax^2 + bx + c$ não são racionais.

9. Uma pessoa possui um gravador de vídeo dotado de um contador que registra o número de voltas dadas pelo carretel da direita. A fita, de 6 horas de duração, está parcialmente gravada. O contador indica 1750 ao final do trecho gravado e 1900 ao final da fita. O problema é saber quanto tempo de gravação ainda está disponível no final da fita.

a) Explique porque não é razoável supor que o tempo de gravação seja proporcional ao número de voltas no contador.

b) Considerando que a fita se enrola em cada carretel segundo círculos concêntricos igualmente espaçados, mostre que o tempo $T(n)$ de gravação após n voltas é dado por uma função da forma $T(n) = an^2 + bn$.

c) Medindo o tempo de gravação corespondente às primeiras 100, 200, 300 e 400 voltas, foram encontrados os dados abaixo. Estes valores são consistentes com o modelo acima?

Volta	Tempo(s)
100	555
200	1176
300	1863
400	2616

d) Quanto tempo de gravação resta na fita?

10. Dado um conjunto de retas do plano, elas determinam um número máximo de regiões quando estão na chamada posição geral: isto é, elas são concorrentes duas a duas e três retas nunca

têm um ponto comum. Seja R_n o número máximo de regiões determinadas por n retas do plano.

a) Quando se adiciona mais uma reta na posição geral a um conjunto de n retas em posição geral, quantas novas regiões são criadas?

b) Deduza de a) que R_n é dada por uma função quadrática de n e obtenha a expressão para R_n.

11. No máximo quantos pontos de interseção existem quando são desenhadas n circunferências?

12. Um estudante anotou a posição, ao longo do tempo, de um móvel sujeito a uma força constante e obteve os dados abaixo:

Instante (seg)	Posição (metros)
0	17
10	45
20	81

Calcule a posição do móvel nos instantes 5 seg, 15 seg e 25 seg.

13. O motorista de um automóvel aplica os freios de modo suave e constante, de modo a imprimir uma força de frenagem constante a seu veículo, até o repouso. O diagrama a seguir mostra a posição do veículo a cada segundo a partir do instante em que os freios foram aplicados.

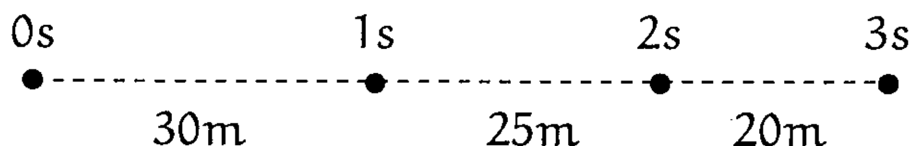


Figura 50

a) Os dados acima são compatíveis com o fato de a força de frenagem ser constante?

b) Qual a posição do veículo 5s após o início da frenagem?

c) Quanto tempo o veículo demora para chegar ao repouso?

d) Qual era a velocidade do veículo no instante em que o motorista começou a aplicar os freios?

14. Um grupo de alunos, ao realizar um experimento no laboratório de Física, fez diversas medidas de um certo comprimento. O instrutor os orientou no sentido de tomar a média aritmética dos valores encontrados como o valor a ser adotado. Este procedimento pode ser justificado do modo abaixo.
Sejam $x_1, x_2, \ldots, x_n$ os valores encontrados. É razoável que o valor adotado x seja escolhido de modo que o erro incorrido pelas diversas medições seja o menor possível. Em geral, este erro é medido através do chamado desvio quadrático total, definido por

$$d(x) = (x - x_1)^2 + (x - x_2)^2 + \cdots + (x - x_n)^2.$$

a) Mostre que $d(x)$ é minimizado quando

$$x = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

b) Suponha agora que se deseje utilizar o desvio absoluto total $e(x) = |x - x_1| + |x - x_1| + \cdots + |x - x_n|$ como medida do erro cometido. Mostre que $e(x)$ é minimizado quando x é a mediana de $x_1, x_2, \ldots, x_n$.

15. Numa vidraçaria há um pedaço de espelho, sob a forma de um triângulo retângulo de lados 60cm, 80cm e 1m. Quer-se, a partir dele, recortar um espelho retangular com a maior área possível. A fim de economizar corte, pelo menos um dos lados do retângulo deve estar sobre um lado do triângulo.

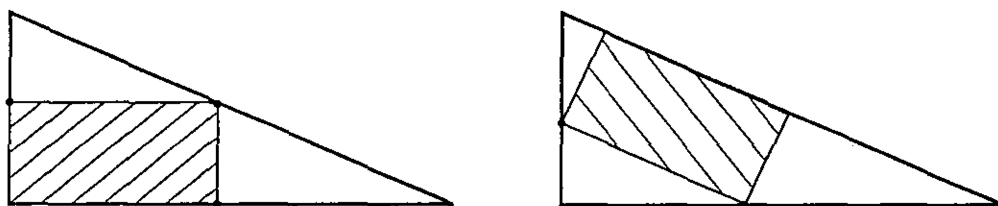


Figura 51

As posições sugeridas são as da figura acima. Em cada caso, determine qual o retângulo de maior área e compare os dois resul-

tados. Discuta se a restrição de um lado estar sobre o contorno do triângulo é realmente necessária para efeito de maximizar a área.

16. Com 80 metros de cerca um fazendeiro deseja circundar uma área retangular junto a um rio para confinar alguns animais.

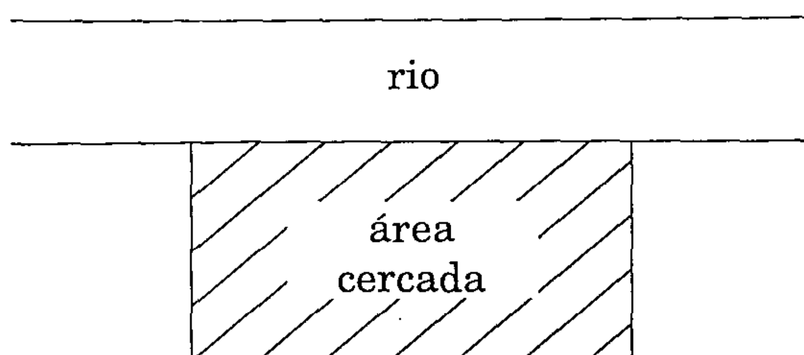


Figura 52

Quais devem ser as medidas do retângulo para que a área cercada seja a maior possível?

17. No instante $t = 0$ o ponto P está em $(-2, 0)$ e o ponto Q em $(0, 0)$. A partir desse instante, Q move-se para cima com velocidade de 1 unidade por segundo e P move-se para a direita com velocidade de 2 unidades por segundo. Qual é o valor da distância mínima entre P e Q?

18. Se x e y são reais tais que $3x + 4y = 12$, determine o valor mínimo de $z = x^2 + y^2$.

19. Um avião de 100 lugares foi fretado para uma excursão. A companhia exigiu de cada passageiro R$ 800,00 mais R$ 10,00 por cada lugar vago. Para que número de passageiros a rentabilidade da empresa é máxima?

20. João tem uma fábrica de sorvetes. Ele vende, em média, 300 caixas de picolés por R$ 20,00. Entretanto, percebeu que, cada vez que diminuía R$ 1,00 no preço da caixa, vendia 40 caixas a mais. Quanto ele deveria cobrar pela caixa para que sua receita fosse máxima?

21. Uma loja está fazendo uma promoção na venda de balas: "Compre x balas e ganhe $x\%$ de desconto". A promoção é válida para compras de até 60 balas, caso em que é concedido o desconto máximo de 60%. Alfredo, Beatriz, Carlos e Daniel compraram 10, 15, 30 e 45 balas, respectivamente. Qual deles poderia ter comprado mais balas e gasto a mesma quantia, se empregasse melhor seus conhecimentos de Matemática?

22. O diretor de uma orquestra percebeu que, com o ingresso a R$ 9,00, em média 300 pessoas assistem aos concertos e que, para cada redução de R$ 1,00 no preço dos ingressos, o público aumenta de 100 espectadores. Qual deve ser o preço do ingresso para que a receita seja máxima?

23. Qual o valor máximo de $21n - n^2$, n inteiro?

24. Faça o gráfico de:

a) $f(x) = |x^2| - |x| + 1$

b) $f(x) = |x^2 - x|$

25. Identifique o conjunto dos pontos (x, y) tais que:

a) $x^2 - 5x + 6 = 0$

b) $y = x^2 - 5x + 6$

26. Resolva a inequação $x^4 + x^2 - 20 > 0$.

27. Determine explicitamente os coeficientes a, b, c do trinômio $f(x) = ax^2 + bx + c$ em função dos valores $f(0)$, $f(1)$ e $f(2)$.

28. Um restaurante a quilo vende 100 kg de comida por dia, a 12 reais o quilo. Uma pesquisa de opinião revelou que, por cada real de aumento no preço, o restaurante perderia 10 clientes, com um consumo médio de 500g cada. Qual deve ser o valor do quilo de comida para que o restaurante tenha a maior receita possível?

29. Um prédio de 1 andar, de forma retangular, com lados proporcionais a 3 e 4, vai ser construído. O imposto predial é de 7 reais por metro quadrado, mais uma taxa fixa de 2.500 reais. A prefeitura concede um desconto de 60 reais por metro linear do perímetro,

como recompensa pela iluminação externa e pela calçada em volta do prédio. Quais devem ser as medidas dos lados para que o imposto seja o mínimo possível? Qual o valor desse imposto mínimo? Esboce o gráfico do valor do imposto como função do lado maior do retângulo.

30. Determine entre os retângulos de mesma área a, aquele que tem o menor perímetro. Existe algum retângulo cujo perímetro seja maior do que os de todos os demais com mesma área?

31. Que forma tem o gráfico da função $f\colon [0,+\infty) \to \mathbb{R}$, dada por $f(x) = \sqrt{x}$?

32. Mostre que a equação $\sqrt{x} + m = x$ possui uma raiz se $m > 0$, duas raízes quando $-\frac{1}{4} < m \leqslant 0$, uma raiz para $m = -1/4$ e nenhuma raiz caso $m < -1/4$.

33. Numa concorrência pública para a construção de uma pista circular de patinação apresentam-se as firmas A e B. A firma A cobra 20 reais por metro quadrado de pavimentação, 15 reais por metro linear do cercado, mais uma taxa fixa de 200 reais para administração. Por sua vez, a firma B cobra 18 reais por metro quadrado de pavimentação, 20 reais por metro linear do cercado e taxa de administração de 600 reais. Para quais valores do diâmetro da pista a firma A é mais vantajosa? Esboce um gráfico que ilustre a situação. Resolva um problema análogo com os números 18, 20 e 400 para A e 20, 10, 150 para B.

34. Dados a, b, c positivos, determinar x e y tais que $xy = c$ e que $ax + by$ seja o menor possível.

35. Cavar um buraco retangular de 1m de largura de modo que o volume cavado seja 300m^3. Sabendo que cada metro quadrado de área cavada custa 10 reais e cada metro de profundidade custa 30 reais, determinar as dimensões do buraco de modo que o seu custo seja mínimo.

36. Dois empresários formam uma sociedade cujo capital é de 100 mil reais. Um deles trabalha na empresa três dias por semana e o outro 2. Após um certo tempo, vendem o negócio e cada um recebe 99 mil reais. Qual foi a contribuição de cada um para formar a sociedade?

37. Nas águas paradas de um lago, Marcelo rema seu barco a 12 km por hora. Num certo rio, com o mesmo barco e as mesmas remadas, ele percorreu 12 km a favor da corrente e 8 km contra a corrente, num tempo total de 2 horas. Qual era a velocidade do rio, quanto tempo ele levou para ir e quanto tempo para voltar?

38. Os alunos de uma turma fizeram uma coleta para juntar 405 reais, custo de uma excursão. Todos contribuiram igualmente. Na última hora, dois alunos desistiram. Com isso, a parte de cada um sofreu um aumento de um real e vinte centavos. Quantos alunos tem a turma?

39. Prove que a função contínua $f: \mathbb{R} \to \mathbb{R}$ é quadrática se, e somente se, para todo $h \in \mathbb{R}$ fixado, a função $\varphi(x) = f(x+h) - f(x)$ é afim e não-constante.

40. Olhando o gráfico da função quadrática $f(x) = x^2$, vê-se que ele parece uma parábola. Se for, quais serão o foco e a diretriz? Por simetria, o foco deve ser $F = (0, t)$ e a diretriz deve ser a reta $y = -t$. Use a definição de parábola para mostrar que $t = 1/4$.

Capítulo 7

Funções Polinomiais

1. Funções Polinomiais vs Polinômios

Diz-se que $p: \mathbb{R} \to \mathbb{R}$ é uma *função polinomial* quando existem números $a_0, a_1, \ldots, a_n$ tais que, para todo $x \in \mathbb{R}$, tem-se

(*) $$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 .$$

Se $a_n \neq 0$, dizemos que p tem *grau* n.

A soma e o produto de funções polinomiais são ainda funções polinomiais. Um exemplo interessante de produto é

$$(x-\alpha)(x^{n-1} + \alpha x^{n-2} + \cdots + \alpha^{n-2}x + \alpha^{n-1}) = x^n - \alpha^n .$$

Dizemos então que $x^n - \alpha^n$ é *divisível* por $x - \alpha$.

Seja p a função polinomial apresentada em (*). Para quaisquer x, α reais, temos

$$p(x) - p(\alpha) = a_n(x^n - \alpha^n) + a_{n-1}(x^{n-1} - \alpha^{n-1}) + \cdots + a_1(x - \alpha).$$

Como cada parcela do segundo membro é divisível por $x - \alpha$, podemos escrever, para todo $x \in \mathbb{R}$:

$$p(x) - p(\alpha) = (x - \alpha)q(x),$$

onde q é uma função polinomial.

Se p tem grau n, q tem grau $n-1$.

Em particular, se α é uma *raiz* de p, isto é, $p(\alpha) = 0$, então

$$p(x) = (x - \alpha)q(x)$$

para todo $x \in \mathbb{R}$. A recíproca é óbvia.

Portanto, α é uma raiz de p se, e somente se, $p(x)$ é divisível por $x-\alpha$. Mais geralmente $\alpha_1, \ldots, \alpha_k$ são raízes de p se, e somente se, para todo $x \in \mathbb{R}$ vale

$$p(x) = (x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_k)q(x),$$

onde q é uma função polinomial de grau $n-k$ se p tem grau n.

Daí resulta que *uma função polinomial de grau n não pode ter mais do que n raízes.*

Uma função polinomial p chama-se *identicamente nula* quando se tem $p(x) = 0$ para todo $x \in \mathbb{R}$. Neste caso, p tem uma infinidade de raízes. (Todo número real é raiz de p.) Então nenhum número natural n é grau de p, a fim de não contradizer o resultado acima. Isto significa que na expressão

$$p(x) = a_n x^n + \cdots + a_1 x + a_0.$$

todos os coeficientes $a_n, a_{n-1} \ldots, a_1, a_0$ são iguais a zero. Concluímos então que a única função polinomial identicamente nula é do tipo

$$0x^n + 0x^{n-1} + \cdots + 0x + 0.$$

Se nos ativermos à letra da definição, a função polinomial identicamente nula não tem grau, pois nenhum dos seus coeficientes é $\neq 0$.

Dadas as funções polinomiais p e q, completando com zeros (se necessário) os coeficientes que faltam, podemos escrevê-las sob as formas

$$p(x) = a_n x^n + \cdots + a_1 x + a_0$$

e

$$q(x) = b_n x^n + \cdots + b_1 x + b_0,$$

sem que isto signifique que ambas têm grau n, pois não estamos dizendo que $a_n \neq 0$ nem que $b_n \neq 0$.

Suponhamos que $p(x) = q(x)$ para todo $x \in \mathbb{R}$, ou seja, que p e q sejam funções iguais. Então a diferença $d = p - q$ é a função

identicamente nula, pois $d(x) = p(x) - q(x) = 0$ para todo $x \in \mathbb{R}$. Mas, para todo $x \in \mathbb{R}$, tem-se

$$d(x) = (a_n - b_n)x^n + \cdots + (a_1 - b_1)x + (a_0 - b_0).$$

Pelo que acabamos de ver sobre funções polinomiais identicamente nulas, segue-se que $a_n - b_n = 0, \ldots, a_1 - b_1 = 0$, $a_0 - b_0 = 0$, ou seja:

$$a_n = b_n, \ldots, a_1 = b_1, \quad a_0 = b_0.$$

Portanto as funções polinomiais p, q assumem o mesmo valor $p(x) = q(x)$ para todo $x \in \mathbb{R}$ se, e somente se, têm os mesmos coeficientes.

Como no caso das funções quadráticas, existe uma diferença sutil entre o conceito de função polinomial e o conceito de polinômio, que apresentaremos agora.

Um *polinômio* é uma expressão formal do tipo

$$p(X) = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0,$$

onde $(a_0, a_1, \ldots, a_n)$ é uma lista ordenada de números reais e X é um símbolo (chamado uma *indeterminada*), sendo X^i uma abreviatura para $X \cdot X \cdots X$ (i fatores). Em essência, o polinômio $p(X)$ é o mesmo que a lista ordenada dos seus coeficientes. Ao escrevê-lo da maneira acima, estamos deixando explícita a intenção de somar e multiplicar polinômios como se fossem funções polinomiais, usando a regra $X^i \cdot X^j = X^{i+j}$. Por definição, os polinômios

$$p(X) = a_n X^n + \cdots + a_1 X + a_0$$

e

$$q(X) = b_n X^n + \cdots + b_1 X + b_0$$

são *iguais* (ou *idênticos*) quando $a_0 = b_0, a_1 = b_1, \ldots, a_n = b_n$.

A cada polinômio

$$p(X) = a_n X^n + \cdots + a_1 X + a_0.$$

faz-se corresponder a função polinomial $\overline{p}: \mathbb{R} \to \mathbb{R}$, definida por $\overline{p}(x) = a_n x^n + \cdots + a_1 x + a_0$, para todo $x \in \mathbb{R}$. Esta correspondência (polinômio) $\mapsto$ (função polinomial) é sobrejetiva, pela própria definição destas funções. A discussão que fizemos acima sobre os coeficientes de funções polinomiais iguais significa que a polinômios distintos correspondem funções polinomiais distintas. Logo, trata-se de uma correspondência biunívoca.

Por esse motivo, não há necessidade de fazer distinção entre o polinômio p e a função polinomial $\overline{p}$. Ambos serão representados pelo mesmo símbolo p e serão chamados indiferentemente de polinômio ou de função polinomial. Além disso, diremos "a função $p(x)$" sempre que não houver perigo de confundí-la com número real que é o valor por ela assumido num certo ponto x.

2. Determinando um Polinômio a Partir de Seus Valores

Um polinômio de grau n é dado quando se conhecem seus $n+1$ coeficientes. Segundo a boa prática matemática, para determinar $n+1$ números é necessário (e muitas vezes suficiente) ter $n+1$ informações. No nosso caso, vale o seguinte resultado:

Dados $n+1$ números reais distintos $x_0, x_1, \ldots, x_n$ e fixados arbitrariamente os valores $y_0, y_1, \ldots, y_n$, existe um, e somente um, polinômio p, de grau $\leqslant n$, tal que

$$p(x_0) = y_0, p(x_1) = y_1, \ldots, p(x_n) = y_n.$$

A parte "somente um" decorre imediatamente do que foi visto na seção anterior pois se p e q são polinômios de grau $\leqslant n$ que assumem os mesmos valores em $n+1$ pontos distintos então a diferença $p - q$ é um polinômio de grau $\leqslant n$ com $n+1$ raízes, logo $p - q = 0$ e $p = q$.

A existência de um polinômio p de grau $\leqslant n$ que assume valores pré-fixados em $n+1$ pontos distintos dados pode ser provada de duas maneiras diferentes. A primeira delas segue as mesmas li-

nhas do caso $n = 2$, já estudado no capítulo anterior, e consiste em resolver o sistema de $n+1$ equações nas $n+1$ incógnitas $a_0, a_1, \ldots, a_n$ abaixo indicado:

$$\begin{aligned} a_n x_0^n + \cdots + a_1 x_0 + a_0 &= y_0 \\ a_n x_1^n + \cdots + a_1 x_1 + a_0 &= y_1 \\ &\vdots \\ a_n x_n^n + \cdots + a_1 x_n + a_0 &= y_n . \end{aligned}$$

Este sistema, no qual as quantidades conhecidas são as potências sucessivas de $x_0, x_1, \ldots, x_n$, tem sempre solução única quando estes $n+1$ números são dois a dois diferentes. [Seu determinante é o determinante de Vandermonde, igual a $\prod_{i<j} (x_i - x_j)$.]

Outra maneira de provar que existe sempre um polinômio de grau $\leqslant n$ que assume nos $n+1$ pontos distintos $x_0, x_1, \ldots, x_n$ os valores arbitrados $y_0, y_1, \ldots, y_n$ consiste em exibir explicitamente esse polinômio, usando a chamada *fórmula de interpolação de Lagrange.*

Apresentamos a seguir os polinômios que resolvem o problema, destacando em especial os casos mais simples, $n = 1$ e $n = 2$.

$n = 1$:

$$p(x) = y_0 \frac{x - x_1}{x_0 - x_1} + y_1 \frac{x - x_0}{x_1 - x_0}.$$

$n = 2$:

$$p(x) = y_0 \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + y_1 \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + y_2 \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}.$$

Caso geral:

$$p(x) = \sum_{i=0}^{n} y_i \cdot \prod_{k \neq i} \left(\frac{x - x_k}{x_i - x_k} \right).$$

Esta é a fórmula de interpolação de Lagrange. Vê-se imediatamente que o polinômio $p(x)$ aí definido cumpre as condições $p(x_0) = y_0$, $p(x_1) = y_1, \ldots, p(x_n) = y_n$. Esse polinômio tem grau

$\leqslant n$ mas seu grau pode perfeitamente ser qualquer número inteiro entre 0 e n.

Por exemplo, se pusermos $x_0 = -1$, $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ e $x_4 = 3$ e procurarmos o polinômio de grau $\leqslant 4$ que assume nesses pontos os valores $-7, 1, 5, 6$ e 25 respectivamente, obteremos

$$p(x) = x^3 - 2x^2 + 5x + 1,$$

que tem grau 3.

E se, dados $n + 1$ pontos distintos, procurarmos o polinômio de grau $\leqslant n$ que se anula em todos esses pontos, a fórmula de Lagrange nos dará o polinômio identicamente nulo, o qual, segundo nossa definição não tem grau. Exceções como esta, e como várias outras que ocorrem quando se estudam polinômios, tornam conveniente atribuir ao polinômio identicamente nulo o grau $-\infty$. (Por exemplo: a convenção gr $0 = -\infty$ torna verdadeira, sem exceções, a afirmação de que o grau do produto de dois polinômios é a soma dos graus dos fatores.)

3. Gráficos de Polinômios

Quando se deseja traçar, ao menos aproximadamente, o gráfico de um polinômio, certas informações de natureza geral são de grande utilidade. Vejamos algumas delas.

Seja $p(x) = a_n x^n + \cdots + a_1 x + a_0$, com $a_n \neq 0$.

Se n é par então, para $|x|$ suficientemente grande, $p(x)$ tem o mesmo sinal de a_n. Este sinal é, portanto, o mesmo, não importando se $x < 0$ ou $x > 0$, desde que $|x|$ seja suficientemente grande.

Se, entretanto, n é ímpar, $p(x)$ tem o mesmo sinal de a_n para valores positivos muito grandes de x e tem o sinal oposto de a_n para valores negativos muito grandes de a_n.

Em ambos os casos (n par ou n ímpar), quando $|x|$ cresce ilimitadamente, $|p(x)|$ também cresce ilimitadamente.

As figuras abaixo esboçam gráficos de polinômios do primeiro, segundo, terceiro e quarto graus. Em cada caso, pode-se dizer logo

qual o sinal do coeficiente do termo de mais alto grau.

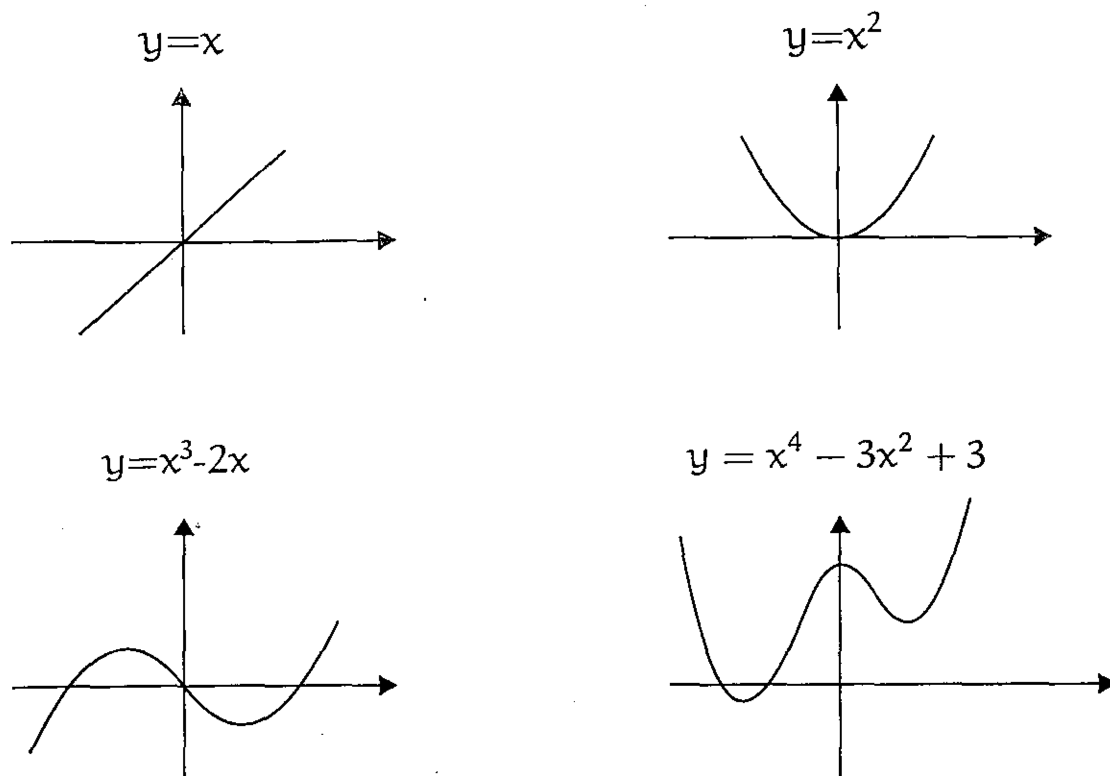


Figura 53

Outra informação útil diz respeito à comparação entre dois polinômios. Se o grau de p é maior do que o grau de q então, para todo x com valor absoluto suficientemente grande, tem-se $|p(x)| > |q(x)|$. Mais ainda: a diferença entre $|p(x)|$ e $|q(x)|$ pode tornar-se tão grande quanto se queira, desde que se tome $|x|$ suficientemente grande.

Um exemplo extremamente simples desta situação ocorre com os polinômios $p(x) = x^2$ e $q(x) = x^6$. Quando $0 < |x| < 1$, x^6 é menor do que x^2 mas, para $|x| > 1$, x^6 supera x^2 e, quando $|x|$ é bastante grande, x^6 é muito, muito maior do que x^2 (ver ilustração na próxima página).

Mais um dado relevante para traçar o gráfico de um polinômio é a localização de suas raízes. É claro que, por motivo da continuidade, se $p(x_1) < 0$ e $p(x_2) > 0$ então p deve possuir uma raiz entre x_1 e x_2. (Esta observação já assegura que todo polinômio de grau ímpar possui ao menos uma raiz real.) Mas como localizar

alguma dessas raízes?

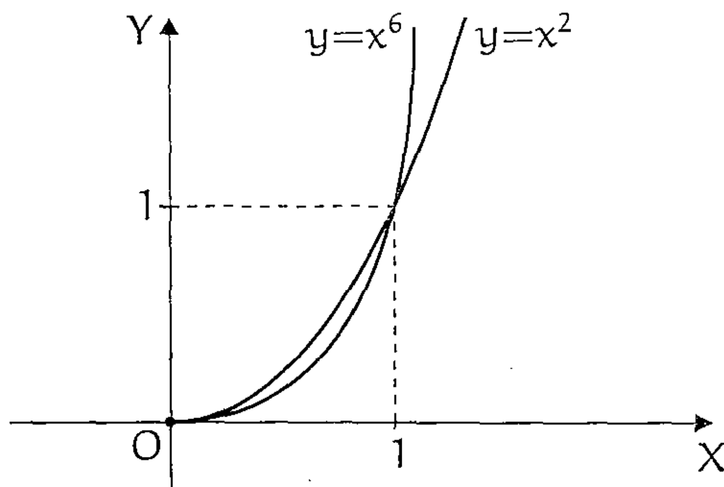


Figura 54

As raízes dos polinômios de grau 2 foram expressas em função dos coeficientes há milênios. Durante a Renascença (meados do século 16) foram obtidas fórmulas para exprimir, mediante radicais, as raízes dos polinômios de terceiro e quarto graus em função dos coeficientes. Na verdade, essas fórmulas têm pouco mais do que mero valor teórico; são demasiadamente complicadas para serem de uso computacional.

Os métodos que se usam atualmente para determinar uma raiz do polinômio p localizada no intervalo $[a, b]$, quando se sabe que $p(a)$ e $p(b)$ têm sinais opostos não se baseiam em fórmulas fechadas, como as que foram obtidas para as equações de grau $\leqslant 4$. Em vez disso, esses métodos se baseiam em *algoritmos aproximativos*, os quais instruem, passo a passo, como proceder para obter uma seqüência de números $x_1, x_2, \ldots, x_n, \ldots$ tais que os valores $p(x_1), p(x_2), \ldots, p(x_n)$ estão cada vez mais próximos de zero.

Um exemplo de algoritmo grandemente eficiente para obter uma raiz da equação $p(x) = 0$ é o *método de Newton*. Segundo este método, se x_1 é um valor próximo de uma raiz, a seqüência $x_1, x_2, \ldots, x_n, \ldots$ de números reais obtidos pela fórmula iterativa

$$x_{n+1} = x_n - \frac{p(x_n)}{p'(x_n)},$$

tem como limite uma raiz de p. Os termos x_n desta seqüência se aproximam bastante rapidamente do limite. Um caso particular do método de Newton já era conhecido pelos babilônios, que calculavam a raiz quadrada de um número positivo a (ou seja, uma raiz da equação $x^2 - a = 0$) tomando um valor inicial x_1 e, a partir dele, construir as aproximações $x_1, x_2, \ldots, x_n, \ldots$ de $\sqrt{a}$ pela fórmula iterativa

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right).$$

Observação: No denominador da fórmula de Newton, $p'(x)$ representa a derivada do polinômio

$$p(x) = a_n x^n + \cdots + a_0,$$

a qual é, por definição,

$$p'(x) = n\,a x^{n-1} + (n-1)a\,x^{n-2} + \cdots + a_1.$$

Exemplo. Mostraremos agora como é eficiente o método de Newton para achar raízes reais de uma equação algébrica. Para isso, consideremos a equação $p(x) = 0$ onde $p(x) = x^5 - 5x^2 + 1$. Então $p'(x) = 5x^4 - 10x$. Começamos observando que $p(1) = -3$ é negativo enquanto que $p(2) = 13$ é positivo, logo deve haver uma raiz real de p entre 1 e 2. Para achar essa raiz, tomamos $x_0 = 2$ como ponto de partida. Obtemos sucessivamente

$$x_1 = x_0 - \frac{p(x_0)}{p'(x_0)} = 2 - \frac{13}{60} = 1,783.$$

$$x_2 = x_1 - \frac{p(x_1)}{p'(x_1)} = 1,783 - \frac{3,124}{32,703} = 1,687.$$

$$x_3 = x_2 - \frac{p(x_2)}{p'(x_2)} = 1,687 - \frac{0,434}{23,627} = 1,668.$$

Com paciência e uma calculadora, poderíamos prosseguir, mas não há necessidade. 1,668 é uma excelente aproximação para a raiz procurada, pois $p(1,668)$ é menor do que 1 milésimo. Uma

aproximação melhor para a raiz procurada seria $1,667977989$, tão próxima do valor que obtivemos que não compensa o esforço de prosseguir o cálculo. De um modo geral, no método de Newton, cada aproximação obtida tem o dobro de dígitos exatos da aproximação anterior. Para mais detalhes teóricos, o leitor pode consultar "Análise Real", vol. 1, pág. 110. E para exercitar-se em contas, notando que $p(0) > 0$ e $p(1) < 0$, pode procurar a raiz de $p(x) = x^5 - 5x^2 + 1$ localizada entre 0 e 1.

Exercícios

1. Sejam $P(x)$ e $p(x)$ polinômios não identicamente nulos, com gr $P(x) \geqslant$ gr $p(x)$. (Onde gr significa o grau do polinômio.) Prove que existe um polinômio $q(x)$ tal que gr $[P(x)-p(x)q(x)] <$ gr $P(x)$. Usando repetidamente este fato, mostre que existem polinômios $q(x)$ e $r(x)$ tais que $P(x) = p(x)q(x) + r(x)$, com gr $r(x) <$ gr $p(x)$. Os polinômios $q(x)$ e $r(x)$, tais que $P(x) = p(x)q(x) + r(x)$ com gr $r(x) <$ gr $p(x)$, chamam-se respectivamente o *quociente* e o *resto* da divisão de $P(x)$ por $p(x)$.

2. Prove a unicidade do quociente e do resto, isto é, se $P(x) = p(x)q_1(x) + r_1(x)$ e $P(x) = p(x)q_2(x) + r_2(x)$, com gr $r_1(x)$ e gr $r_2(x)$ ambos menores do que gr $p(x)$, então $q_1(x) = q_2(x)$ e $r_1(x) = r_2(x)$ para todo $x \in \mathbb{R}$.

3. Diz-se que o número real α é uma raiz de *multiplicidade* m do polinômio $p(x)$ quando se tem $p(x) = (x - \alpha)^m q(x)$, com $q(\alpha) \neq 0$. (Se $m = 1$ ou $m = 2$, α chama-se respectivamente uma *raiz simples* ou uma *raiz dupla*.) Prove que α é uma raiz simples de $p(x)$ se, e somente se, tem-se $p(\alpha) = 0$ e $p'(\alpha) \neq 0$. Prove também que α é uma raiz dupla de $p(x)$ se, e somente se, $p(\alpha) = p'(\alpha) = 0$ e $p''(\alpha) \neq 0$, Generalize.

4. Certo ou errado: α é raiz dupla de $p(x)$ se, e somente se, é raiz simples de $p'(x)$.

5. Determine o poliômio $p(x)$ de menor grau possível tal que

$p(1) = 2$, $p(2) = 1$, $p(3) = 4$ e $p(4) = 3$.

6. Seja $p(x)$ um polinômio cujo grau n é um número ímpar. Mostre que existem números reais x_1, x_2 tais que $p(x_1) < 0$ e $p(x_2) > 0$. Conclua daí que todo polinômio de grau ímpar admite pelo menos uma raiz real.

7. Mostre que se n é um número par então o polinômio $p(x) = x^n + x^{n-1} + \cdots + x + 1$ não possui raiz real.

8. Tomando $x_0 = 3$, use a relação de recorrência

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{5}{x_n}\right)$$

para calcular $\sqrt{5}$ com três algarismos decimais exatos. (Por exemplo: sabemos que 1,414 é uma aproximação de $\sqrt{2}$ com três algarismos decimais exatos porque $1,414^2 < 2 < 1,415^2$.)

9. Usando o método de Newton, estabeleça um processo iterativo para calcular $\sqrt[3]{a}$ e aplique-o a fim de obter um valor aproximado de $\sqrt[3]{2}$.

Capítulo 8

Funções Exponenciais e Logarítmicas

1. Introdução

Vimos no Capítulo 5 que se $f: \mathbb{R} \to \mathbb{R}$ é uma função afim então o acréscimo $f(x+h) - f(x)$, sofrido por f quando se passa de x para $x+h$, depende apenas do acréscimo h dado a x mas não depende do próprio valor de x. Isto é óbvio, uma vez que $f(x) = ax + b$ implica $f(x+h) - f(x) = ah$. O mais importante, tendo em vista as aplicações, é que quando f é monótona crescente, ou decrescente, vale a recíproca: se $f(x+h) - f(x)$ não depende de x então f é afim.

O Exemplo 8 do Capítulo 5 dizia respeito a uma quantia x, investida durante um prazo fixo e determinado, gerando no final desse período o valor $f(x)$. Constatou-se alí que $f(x)$ é uma função linear de x.

Neste capítulo, consideraremos uma quantia c_0, aplicada a juros fixos, capitalizados continuamente. Se chamarmos de $c(t)$ o capital gerado a partir daquela quantia inicial depois de decorrido o tempo t, é claro que $c(t)$ é uma função crescente de t.

Notamos ainda que se $t < t'$ então o acréscimo $c(t'+h) - c(t')$, experimentado pelo capital após o decurso do tempo h, a partir do momento t', é maior do que o rendimento $c(t+h) - c(t)$ depois de decorrido o mesmo tempo h, a partir do momento anterior t, pois o capital acumulado $c(t')$, sendo maior do que $c(t)$, deve produzir maior renda.

Assim, $c(t)$ não é uma função afim de t, já que $c(t+h) -$

$c(t)$ depende não apenas de h mas de t também. Esta conclusão negativa indica que se deve buscar outro instrumento matemático, diferente da função afim, para modelar a presente situação.

Analisando este problema mais detidamente, vemos que a diferença $c(t+h) - c(t)$ pode ser considerada como o lucro obtido quando se investiu a quantia $c(t)$ durante o prazo h. Portanto, como vimos acima, $c(t+h) - c(t)$ deve ser proporcional à quantia aplicada $c(t)$, ou seja, $c(t+h) - c(t) = \varphi \cdot c(t)$, onde o fator de proporcionalidade $\varphi = \varphi(h)$ depende evidentemente do prazo h. A afirmação de que $\varphi(h) = [c(t+h) - c(t)]/c(t)$ não depende de t é a expressão matemática do fato de que os juros são fixos. Como $[c(t+h) - c(t)]/c(t) = [c(t+h)/c(t)] - 1$, esta afirmação equivale a dizer que o quociente $c(t+h)/c(t)$ não depende de t.

Portanto, quando os juros são fixos, se $c(t_1+h)/c(t_1) = 2$, por exemplo, então $c(t_2+h)/c(t_2) = 2$ para qualquer t_2 (e o mesmo h). Isto quer dizer que o tempo h necessário para que um capital seja dobrado é o mesmo em todas as ocasiões e para qualquer valor desse capital, pequeno ou grande.

Vemos então que o modelo matemático conveniente para descrever a variação de um capital aplicado a juros fixos, em função do tempo, deve ser uma função crescente $c(t)$ tal que o acréscimo relativo $[c(t+h) - c(t)]/c(t)$ dependa apenas de h mas não de t.

Conforme será estabelecido neste capítulo, as únicas funções com estas propriedades são as da forma $c(t) = c_0 \cdot a^t$.

Uma situação análoga ocorre quando se estuda a desintegração radioativa. Os átomos de uma substância radioativa (como o rádio e o urânio, por exemplo) tendem a se desintegrar, emitindo partículas e transformando-se noutra substância. As partículas emitidas não alteram consideravelmente a massa total do corpo mas, com o passar do tempo, a quantidade da substância original diminui (aumentando, conseqüentemente, a massa da nova substância transformada). Isto ocorre de tal modo que, em cada instante, a quantidade de matéria que se está desintegrando naquele momento é proporcional à massa da substância original que

ainda resta.

Assim sendo, se chamarmos (como fazem os cientistas) de *meia-vida* de uma substância radioativa o tempo necessário para que se desintegre a metade da massa de um corpo formado por aquela substância, constatamos que a meia-vida é um número intrinsecamente associado a cada substância radioativa: o tempo necessário para reduzir à metade a radioatividade de uma tonelada de urânio é igual ao tempo que leva um grama da mesma substância para ter sua metade desintegrada.

A propósito: os vários isótopos do urânio têm meia-vida da ordem de 10^9 anos. Enquanto isso, a meia-vida do rádio 224 é de 3 dias e 15 horas.

De um modo geral, se designarmos por $m = m(t)$ a massa da substância radioativa presente no corpo no instante t, veremos que m é uma função decrescente de t e, além disso, a perda relativa $[m(t+h)-m(t)]/m(t)$, ocorrida após o decurso do tempo h, depende apenas de h mas não do instante inicial t, ou seja, da massa $m(t)$ existente naquela ocasião.

Outra vez constatamos a necessidade de uma função real de variável real $m\colon \mathbb{R} \to \mathbb{R}$, que seja monótona (desta vez, decrescente) e tal que a variação relativa $[m(t+h) - m(t)]/m(t)$ dependa apenas de h. Ou, equivalentemente, que a razão $m(t+h)/m(t)$ não dependa de t mas somente de h.

Mostraremos neste capítulo que as únicas funções com essas propriedades são as do tipo $m(t) = b \cdot a^t$ (com $0 < a < 1$). Os exemplos que acabamos de mencionar ilustram algumas das inúmeras situações em que ocorrem as funções do tipo exponencial, que estudaremos agora.

Começaremos nosso estudo com uma revisão das potências com expoente racional.

2. Potências de Expoente Racional

Seja a um número real positivo. Para todo $n \in \mathbb{N}$, a potência a^n, de base a e expoente n é definida como o produto de n fatores iguais a

a. Para $n = 1$, como não há produto de um só fator, põe-se $a^1 = a$, por definição.

A definição indutiva de a^n é: $a^1 = a$ e $a^{n+1} = a \cdot a^n$.

Para quaisquer $m, n \in \mathbb{N}$ tem-se

$$a^m \cdot a^n = a^{m+n}$$

pois em ambos os membros desta igualdade temos o produto de $m + n$ fatores iguais a a. Segue-se que, para $m_1, m_2, \ldots, m_k$ quaisquer, vale

$$a^{m_1} \cdot a^{m_2} \cdot \ldots \cdot a^{m_k} = a^{m_1+m_2+\cdots+m_k}.$$

Em particular, se $m_1 = \cdots = m_k = m$, vem $(a^m)^k = a^{mk}$.

Se $a > 1$ então, multiplicando ambos os membros desta desigualdade por a^n, obtemos $a^{n+1} > a^n$. Portanto,

$$a > 1 \quad \Rightarrow \quad 1 < a < a^2 < \cdots < a^n < a^{n+1} < \cdots .$$

Além disso,

$$0 < a < 1 \quad \Rightarrow \quad 1 > a > a^2 > \cdots > a^n > a^{n+1} > \cdots$$

como se vê multiplicando ambos os membros da desigualdade $a < 1$ pelo número positivo a^n.

Portanto a seqüência cujo n-ésimo termo é a^n é crescente quando $a > 1$ e decrescente se $0 < a < 1$. Para $a = 1$, esta seqüência é constante, com todos os seus termos iguais a 1.

Existem seqüências crescentes que são limitadas superiormente. Um exemplo disso é

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots, \frac{n}{n+1}, \ldots$$

onde se tem

$$\frac{n}{n+1} < 1$$

para todo $n \in \mathbb{N}$.

Entretanto, *se* $a > 1$, *a seqüência formada pelas potências* a^n, $n \in \mathbb{N}$, *é ilimitada superiormente: nenhum número real c, por maior que seja, pode ser superior a todas as potências* a^n. Noutras

palavras, dado arbitrariamente $c \in \mathbb{R}$, pode-se sempre achar $n \in \mathbb{N}$ tal que $a^n > c$.

Para provar isto, escrevemos $a = 1 + d$, $c > 0$. Pela desigualdade de Bernoulli, temos $a^n > 1 + nd$. Logo, se tomarmos $n > (c-1)/d$, teremos $1 + nd > c$ e, com maior razão, $a^n > c$.

Exemplo. Seja $a = 1{,}000001$ (um inteiro e um milionésimo). As potências sucessivas $a, a^2, a^3, \ldots$, a princípio próximas de 1, podem tornar-se tão grandes quanto se deseje, desde que o expoente seja tomado suficientemente grande. Se usarmos o argumento acima para obter uma potência de a que seja superior a 1 bilhão, devemos tomar um expoente da ordem de 10^{14}. Na realidade, usando uma calculadora, vemos que para ter $(1{,}000001)^n >$ um bilhão basta tomar $n > 21$ milhões. É que, ao demonstrarmos que as potências sucessivas de um número maior do que 1 crescem acima de qualquer limite prefixado, nos preocupamos mais em usar um raciocínio simples e claro do que obter o menor expoente possível.

Para exprimir que a seqüência crescente (a^n) é ilimitada superiormente, escrevemos $\lim\limits_{n\to\infty} a^n = +\infty$ e dizemos que a^n tende ao infinito quando n cresce indefinidamente (supondo $a > 1$!).

De modo análogo, se $0 < a < 1$ então as potências sucessivas $a, a^2, a^3, \ldots$ decrescem abaixo de qualquer cota positiva: fixado arbitrariamente um número $c > 0$, por menor que seja, pode-se sempre achar um expoente $n \in \mathbb{N}$ tal que $a^n < c$.

Com efeito, sendo $0 < a < 1$, se escrevermos $b = 1/a$, teremos $b > 1$. Logo, pelo que acabamos de ver, podemos achar $n \in \mathbb{N}$ tal que $b^n > 1/c$, ou seja, $\frac{1}{a^n} > \frac{1}{c}$, donde $a^n < c$.

Este resultado significa que $\lim\limits_{n\to\infty} a^n = 0$ quando $0 < a < 1$. (A expressão $\lim\limits_{n\to\infty} a^n = 0$ lê-se "o limite de a^n, quando n tende ao infinito, é igual a zero".)

Procuremos agora atribuir um significado à potência a^n, quando $n \in \mathbb{Z}$ é um número inteiro, que pode ser negativo ou zero. Isto deve ser feito de modo que seja mantida a regra fundamental

$a^m \cdot a^n = a^{m+n}$.

Em primeiro lugar, qual deve ser o valor de a^0?

Como a igualdade $a^0 \cdot a^1 = a^{0+1}$ deve ser válida, teremos $a^0 \cdot a = a$, logo a única definição possível é $a^0 = 1$.

Em seguida, dado qualquer $n \in \mathbb{N}$, devemos ter

$$a^{-n} \cdot a^n = a^{-n+n} = a^0 = 1, \quad \text{logo} \quad a^{-n} = \frac{1}{a^n}.$$

Assim, se quisermos estender o conceito de potência do número real $a > 0$, para admitir expoentes inteiros quaisquer e preservar a igualdade $a^m \cdot a^n = a^{m+n}$, a única definição possível consiste em pôr $a^0 = 1$ e $a^{-n} = 1/a^n$ para todo $n \in \mathbb{N}$.

A função $f: \mathbb{Z} \to \mathbb{R}^+$, dada por $f(n) = a^n$, $n \in \mathbb{Z}$, além de cumprir a igualdade fundamental

$$f(m+n) = f(m) \cdot f(n),$$

é ainda crescente quando $a > 1$ e decrescente quando $0 < a < 1$. Segue-se, em particular que, para $a > 1$ e $n \in \mathbb{N}$, tem-se $a^{-n} < 1 < a^n$ e, para $0 < a < 1$, tem-se $a^n < 1 < a^{-n}$ pois $-n < 0 < n$ e $a^0 = 1$.

De $a^m \cdot a^n = a^{m+n}$ segue-se que $(a^m)^n = a^{mn}$ ainda quando $m, n \in \mathbb{Z}$.

Prosseguindo, vejamos que sentido pode ser dado à potência a^r quando $r = m/n$ é um número racional (onde $m \in \mathbb{Z}$ e $n \in \mathbb{N}$), de modo que continue válida a regra $a^r \cdot a^s = a^{r+s}$. Desta igualdade resulta, que se deve ter, para $r = m/n$:

$$(a^r)^n = a^r \cdot a^r \cdot \ldots \cdot a^r = a^{r+r+\cdots+r} = a^{rn} = a^m.$$

Portanto a^r é o número real positivo cuja n-ésima potência é igual a a^m. Por definição de raiz, este número é $\sqrt[n]{a^m}$, a raiz n-ésima de a^m. Assim, a única maneira de definir a potência a^r, com $r = m/n$, $m \in \mathbb{Z}$, $n \in \mathbb{N}$, consiste em pôr

$$a^{m/n} = \sqrt[n]{a^m}.$$

Depois de dar esta definição, há alguns detalhes que devem ser examinados. Em primeiro lugar, como se tem $m/n = mp/np$ para todo $p \in \mathbb{N}$, é preciso mostrar que $\sqrt[n]{a^m} = \sqrt[np]{a^{mp}}$ a fim de que a definição não seja ambígua. Em segundo lugar, deve-se mostrar que a definição dada assegura a validez da regra $a^r \cdot a^s = a^{r+s}$ para $r, s \in \mathbb{Q}$. E finalmente, cumpre provar que a função $f: \mathbb{Q} \to \mathbb{R}^+$, definida por $f(r) = a^r$, é crescente quando $a > 1$ e decrescente quando $0 < a < 1$. Esses pormenores estão tratados no Capítulo 2 do livro "Logaritmos", da Coleção do Professor de Matemática.

A função $f: \mathbb{Q} \to \mathbb{R}^+$, definida por $f(r) = a^r$, não é sobrejetiva. Noutras palavras, fixado $a > 0$, nem todo número real positivo é da forma a^r com r racional. Isto fica evidente se observarmos que, como $\mathbb{Q}$ é um conjunto enumerável, o mesmo deve ocorrer com sua imagem $f(\mathbb{Q})$, porém $\mathbb{R}^+$ não é enumerável. De um modo mais elementar, este fato pode ser ilustrado mediante um exemplo. Tomemos $a = 10$ e indaguemos se existe algum número racional $r = m/n$ tal que $10^{m/n} = 11$, ou seja, tal que $10^m = 11^n$, onde $m, n \in \mathbb{N}$. É claro que, para qualquer $m \in \mathbb{N}$, 10^m se escreve como 1 seguido de m zeros enquanto 11^n não pode ter esta forma. Logo o número real positivo 11 não pertence à imagem da função $r \mapsto 10^r$, de $\mathbb{Q}$ em $\mathbb{R}^+$.

As potências a^r, com expoente racional, embora não contenham todos os números reais positivos, estão espalhadas por toda parte em $\mathbb{R}^+$, desde que seja $a \neq 1$. Este é o conteúdo do lema abaixo. A demonstração do mesmo, embora elementar, é um tanto técnica e pode ser omitida numa primeira leitura.

Lema: *Fixado o número real positivo $a \neq 1$, em todo intervalo de $\mathbb{R}^+$ existe alguma potência a^r, com $r \in \mathbb{Q}$.*

Demonstração: Dados $0 < \alpha < \beta$, devemos achar $r \in \mathbb{Q}$ tal que a potência a^r pertença ao intervalo $[\alpha, \beta]$, isto é, $\alpha \leqslant a^r \leqslant \beta$. Por simplicidade, suporemos a e α maiores do que 1. Os demais casos podem ser tratados de modo análogo. Como as potências de expoente natural de números maiores do que 1 crescem acima de

qualquer cota prefixada, podemos obter números naturais M e n tais que

$$\alpha < \beta < a^M \qquad \text{e} \qquad 1 < a < \left(1 + \frac{\beta - \alpha}{a^M}\right)^n.$$

Da última relação decorrem sucessivamente

$$1 < a^{1/n} < 1 + \frac{\beta - \alpha}{a^M} \qquad \text{e} \qquad 0 < a^M(a^{1/n} - 1) < \beta - \alpha.$$

Logo

$$\frac{m}{n} \leqslant M \Rightarrow 0 < a^{\frac{m}{n}}(a^{\frac{1}{n}} - 1) < \beta - \alpha \quad \Leftrightarrow \quad 0 < a^{\frac{m+1}{n}} - a^{\frac{m}{n}} < \beta - \alpha.$$

Assim, as potências

$$a^0 = 1, \quad a^{1/n}, \quad a^{2/n}, \ldots, \quad a^M$$

são extremos de intervalos consecutivos, todos de comprimento menor do que o comprimento $\beta - \alpha$ do intervalo $[\alpha, \beta]$. Como $[\alpha, \beta] \subset [1, a^M]$, pelo menos um desses extremos, digamos $a^{\frac{m}{n}}$, está contido no intervalo $[\alpha, \beta]$.

3. A Função Exponencial

Seja a um número real positivo, que suporemos sempre diferente de 1. A *função exponencial de base* a, $f: \mathbb{R} \to \mathbb{R}^+$, indicada pela notação $f(x) = a^x$, deve ser definida de modo a ter as seguintes propriedades, para quaisquer $x, y \in \mathbb{R}$:

1) $a^x \cdot a^y = a^{x+y}$;
2) $a^1 = a$;
3) $x < y \Rightarrow a^x < a^y$ quando $a > 1$ e
   $x < y \Rightarrow a^y < a^x$ quando $0 < a < 1$.

É interessante observar que se uma função $f: \mathbb{R} \to \mathbb{R}$ tem a propriedade 1) acima, isto é, $f(x+y) = f(x) \cdot f(y)$, então f não pode assumir o valor 0, a menos que seja identicamente nula. Com efeito, se existir algum $x_0 \in \mathbb{R}$ tal que $f(x_0) = 0$ então, para todo $x \in \mathbb{R}$ teremos

$$f(x) = f(x_0 + (x - x_0)) = f(x_0) \cdot f(x - x_0) = 0 \cdot f(x - x_0) = 0,$$

logo f será identicamente nula.

Mais ainda: se $f\colon \mathbb{R} \to \mathbb{R}$ tem a propriedade 1) e não é identicamente nula então $f(x) > 0$ para todo $x \in \mathbb{R}$, pois

$$f(x) = f\left(\frac{x}{2} + \frac{x}{2}\right) = f\left(\frac{x}{2}\right) \cdot f\left(\frac{x}{2}\right) = \left[f\left(\frac{x}{2}\right)\right]^2 > 0.$$

Assim, diante da propriedade 1), tanto faz dizer que o contra-domínio de f é $\mathbb{R}$ como dizer que é $\mathbb{R}^+$. A vantagem de tomar $\mathbb{R}^+$ como contra-domínio é que se terá f sobrejetiva, como veremos.

Se uma função $f\colon \mathbb{R} \to \mathbb{R}$ tem as propriedades 1) e 2) então, para todo $n \in \mathbb{N}$ tem-se

$$f(n) = f(1 + 1 + \cdots + 1) = f(1) \cdot f(1) \cdot \ldots \cdot f(1) = a \cdot a \cdot \ldots \cdot a = a^n.$$

Usando a propriedade 1), resulta daí, como mostramos na seção anterior, que, para todo número racional $r = m/n$, com $n \in \mathbb{N}$, deve-se ter $f(r) = a^r = \sqrt[n]{a^m}$.

Portanto $f(r) = a^r$ é a única função $f\colon \mathbb{Q} \to \mathbb{R}^+$ tal que $f(r+s) = f(r) \cdot f(s)$ para quaisquer $r, s \in \mathbb{Q}$ e $f(1) = a$.

A propriedade 3) diz que a função exponencial deve ser crescente quando $a > 1$ e decrescente quando $0 < a < 1$.

Daí resultará que existe uma única maneira de definir o valor $f(x) = a^x$ quando x é irracional. Para fixar as idéias, suporemos $a > 1$. Então a^x tem a seguinte propriedade:

$$r < x < s, \text{ com } r, s \in \mathbb{Q} \quad \Rightarrow \quad a^r < a^x < a^s.$$

Ou seja, a^x é o número real cujas aproximações por falta são a^r, com $r < x$, $r \in \mathbb{Q}$, e cujas aproximações por excesso são a^s, com $x < s$, $s \in \mathbb{Q}$. Não podem existir dois números reais diferentes, digamos $A < B$, com a propriedade acima. Se existissem tais A e B teríamos

$$r < x < s, \ r, s \in \mathbb{Q} \quad \Rightarrow \quad a^r < A < B < a^s$$

e então o intervalo $[A, B]$ não conteria nenhuma potência de a com expoente racional, contrariando o Lema da seção anterior.

Portanto, quando x é irracional, a^x é o (único) número real cujas aproximações por falta são as potências a^r, com r racional menor do que x e cujas aproximações por excesso são as potências a^s, com s racional maior do que x.

Definindo a^x para todo $x \in \mathbb{R}$, não há maiores dificuldades para verificar que, de fato, são válidas as propriedades 1), 2) e 3) acima enunciadas. Além disso, tem-se ainda

4) *A função* $f: \mathbb{R} \to \mathbb{R}^+$*, definida por* $f(x) = a^x$*, é ilimitada superiormente.*

Com efeito, todo intervalo em $\mathbb{R}^+$ contém valores $f(r) = a^r$ segundo o Lema da seção anterior.

Mais precisamente: se $a > 1$ então a^x cresce sem limites quando $x > 0$ é muito grande. E se $0 < a < 1$ então a^x torna-se arbitrariamente grande quando $x < 0$ tem valor absoluto grande.

5) *A função exponencial é contínua.*

Isto significa que, dado $x_0 \in \mathbb{R}$, é possível tornar a diferença $|a^x - a^{x_0}|$ tão pequena quanto se deseje, desde que x seja tomado suficientemente próximo de x_0. Dito de outro modo: o limite de a^x quando x tende a x_0 é igual a a^{x_0}. Em símbolos: $\lim_{x \to x_0} a^x = a^{x_0}$.

Esta afirmação pode ser provada assim: escrevemos $x = x_0 + h$, logo $x - x_0 = h$ e então $|a^x - a^{x_0}| = a^{x_0}|a^h - 1|$. Ora, sabemos que a^h pode ser tornado tão próximo de 1 quanto desejemos, desde que tomemos h suficientemente pequeno. Como a^{x_0} é constante, podemos fazer o produto $a^{x_0}|a^h - 1|$ tão pequeno quanto o queiramos, logo $\lim_{x \to x_0} |a^x - a^{x_0}| = 0$, ou seja, $\lim_{x \to x_0} a^x = a^{x_0}$.

6) *A função exponencial* $f: \mathbb{R} \to \mathbb{R}^+$, $f(x) = a^x$, $a \neq 1$, *é sobrejetiva.*

Esta afirmação quer dizer que para todo número real $b > 0$ existe algum $x \in \mathbb{R}$ tal que $a^x = b$. (Todo número real positivo é uma potência de a.) Para prová-la, usamos o Lema da seção anterior e escolhemos, para cada $n \in \mathbb{N}$, uma potência a^{r_n}, com $r_n \in \mathbb{Q}$, no intervalo $\left(b - \frac{1}{n}, b + \frac{1}{n}\right)$, de modo que $|b - a^{r_n}| < 1/n$ portanto $\lim_{x \to x_0} a^{r_n} = b$. Para fixar as idéias, supomos $a > 1$. Escolhemos as

potências a^{r_n} sucessivamente, tais que

$$a^{r_1} < a^{r_2} < \cdots < a^{r_n} < \cdots < b.$$

Certamente, podemos fixar $s \in \mathbb{Q}$ tal que $b < a^s$. Então a monotonicidade da função a^x nos assegura que $r_1 < r_2 < \cdots < r_n < \cdots < s$.

Assim, (r_n) é uma seqüência monótona, limitada superiormente por s. A completeza de $\mathbb{R}$ garante então que os r_n são valores aproximados por falta de um número real x, ou seja, $\lim_{x \to x_0} r_n = x$. A função exponencial sendo contínua, temos então $a^x = \lim_{x \to x_0} a^{r_n} = b$, como queríamos demonstrar.

Vemos, pois, que para todo número real positivo a, diferente de 1, a função exponencial $f : \mathbb{R} \to \mathbb{R}^+$, dada por $f(x) = a^x$, é uma correspondência biunívoca entre $\mathbb{R}$ e $\mathbb{R}^+$, crescente se $a > 1$, decrescente se $0 < a < 1$, com a propriedade adicional de transformar somas em produtos, isto é, $f(x + y) = f(x) \cdot f(y)$.

(A injetividade da função $x \mapsto a^x$ decorre da sua monotonicidade. Se $a > 1$, por exemplo, então

$$x > y \Rightarrow a^x > a^y$$

e

$$x < y \Rightarrow a^x < a^y,$$

portanto $x \neq y \Rightarrow a^x \neq a^y$.)

Tem-se ainda

$$\lim_{x \to +\infty} a^x = +\infty \quad \text{se} \quad a > 1,$$

$$\lim_{x \to +\infty} a^x = 0 \quad \text{se} \quad 0 < a < 1,$$

$$\lim_{x \to -\infty} a^x = 0 \quad \text{se} \quad a > 1 \quad \text{e}$$

$$\lim_{x \to -\infty} a^x = +\infty \quad \text{se} \quad 0 < a < 1.$$

A figura exibe o gráfico de $f(x) = a^x$ nos casos $a > 1$ e $0 < a < 1$

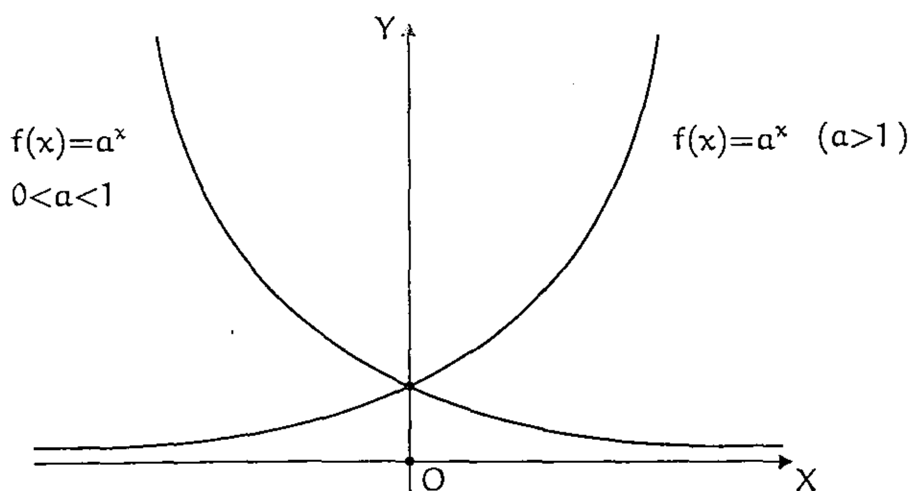


Figura 55

Quando $a > 1$, nota-se que, quando x varia da esquerda para a direita, a curva exponencial $y = a^x$ apresenta um crescimento bastante lento enquanto x é negativo. À medida que x cresce, o crescimento de y se torna cada vez mais acelerado. Isto se reflete na inclinação da tangente ao gráfico; para valores positivos muito grande de x, a tangente é quase vertical. O crescimento exponencial supera o de qualquer polinômio. Se compararmos o gráfico de $y = 2^x$ (por exemplo) com o de $y = x^{10}$, veremos que, para $0 < x < 1{,}077$ temos $x^{10} < 2^x$. Para $1{,}077 < x < 58{,}77$ tem-se $x^{10} > 2^x$ e, para todo $x > 58{,}77$ tem-se sempre $2^x > x^{10}$.

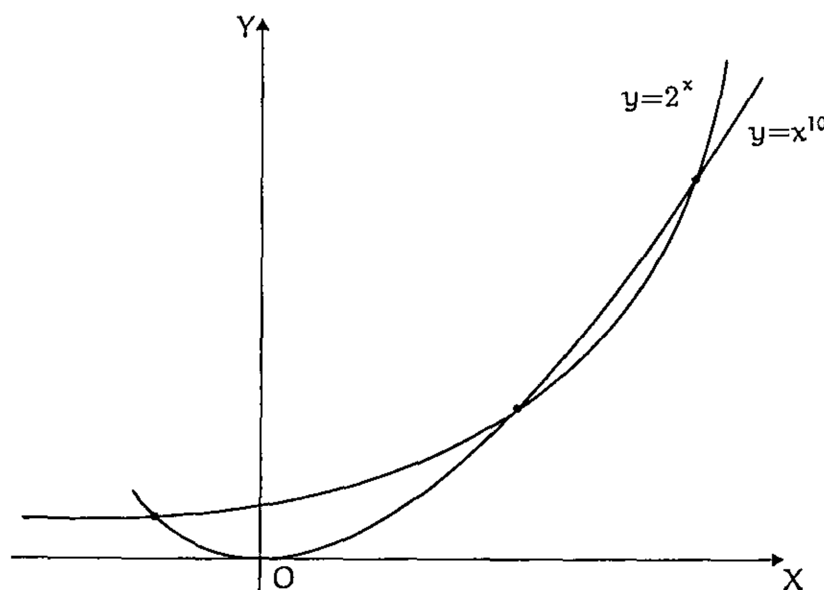


Figura 56

4. Caracterização da Função Exponencial

As funções exponenciais são, juntamente com as funções afins e as quadráticas, os modelos matemáticos mais utilizados para resolver problemas elementares. As funções afins ocorrem em praticamente todos os problemas durante os oito primeiros anos da escola e, com menos exclusividade, porém ainda com grande destaque, nos três anos finais. Por sua vez, as funções quadráticas e exponenciais aparecem nesses três últimos anos, embora tenham, principalmente as últimas, importância considerável na universidade, bem como nas aplicações de Matemática em atividades científicas ou profissionais.

Uma vez decidido que o modelo adequado para um determinado problema é uma função afim, quadrática ou exponencial, a partir daí o tratamento matemático da questão não oferece maiores dificuldades. As dúvidas que possam surgir acontecem geralmente, antes, na escolha do instrumento matemático apropriado para o problema que se estuda. Para que essa escolha possa ser feita corretamente, é preciso saber quais são as propriedades características de cada tipo de função. Nos Capítulos 5 e 6, vimos propriedades que caracterizam as funções afins e quadráticas. Vamos agora fazer o mesmo com as funções exponenciais.

Teorema: (Caracterização da função exponencial.) *Seja* $f\colon \mathbb{R} \to \mathbb{R}^+$ *uma função monótona injetiva (isto é, crescente ou decrescente). As seguintes afirmações são equivalentes:*

(1) $f(nx) = f(x)^n$ *para todo* $n \in \mathbb{Z}$ *e todo* $x \in \mathbb{R}$;

(2) $f(x) = a^x$ *para todo* $x \in \mathbb{R}$, *onde* $a = f(1)$;

(3) $f(x+y) = f(x) \cdot f(y)$ *para quaisquer* $x, y \in \mathbb{R}$.

Demonstração: Provaremos as implicações $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1)$. A fim de mostrar que $(1) \Rightarrow (2)$ observamos inicialmente que a hipótese (1) acarreta que, para todo número racional $r = m/n$ (com $m \in \mathbb{Z}$ e $n \in \mathbb{N}$) tem-se $f(rx) = f(x)^r$. Com efeito, como $nr = m$,

podemos escrever

$$f(rx)^n = f(nrx) = f(mx) = f(x)^m,$$

logo $f(rx) = f(x)^{m/n} = f(x)^r$.

Assim, se pusermos $f(1) = a$, teremos $f(r) = f(r \cdot 1) = f(1)^r = a^r$ para todo $r \in \mathbb{Q}$. Para completar a demonstração de que (1) $\Rightarrow$ (2) suponhamos, a fim de fixar as idéias que f seja crescente, logo $1 = f(0) < f(1) = a$. Admitamos, por absurdo, que exista um $x \in \mathbb{R}$ tal que $f(x) \neq a^x$. Digamos, por exemplo, que seja $f(x) < a^x$. (O caso $f(x) > a^x$ seria tratado analogamente.) Então, pelo Lema da seção 2, existe um número racional r tal que $f(x) < a^r < a^x$, ou seja, $f(x) < f(r) < a^x$. Como f é crescente, tendo $f(x) < f(r)$ concluímos que $x < r$. Por outro lado, temos também $a^r < a^x$, logo $r < x$. Esta contradição completa a prova de que (1) $\Rightarrow$ (2). As implicações restantes, (2) $\Rightarrow$ (3) e (3) $\Rightarrow$ (1) são óbvias.

Observação. O Teorema de caracterização pode ser enunciado de um modo ligeiramente diferente, substituindo a hipótese de monotonicidade pela suposição de que f seja contínua. A demonstração do passo (1) $\Rightarrow$ (2) muda apenas no caso x irracional. Então tem-se $x = \lim\limits_{n\to\infty} r_n = r_n$, $r_n \in \mathbb{Q}$, logo, pela continuidade de f, deve ser

$$f(x) = \lim_{n\to\infty} f(r_n) = \lim_{n\to\infty} a^{r_n} = a^x.$$

Dizemos que uma função $g: \mathbb{R} \to \mathbb{R}$ é de *tipo exponencial* quando se tem $g(x) = ba^x$ para todo $x \in \mathbb{R}$, onde a e b são constantes positivas. Se $a > 1$, g é crescente e se $0 < a < 1$ g é decrescente.

Se a função $g: \mathbb{R} \to \mathbb{R}$ é de tipo exponencial então, para quaisquer $x, h \in \mathbb{R}$, os quocientes

$$\frac{g(x+h) - g(x)}{g(x)} = a^h - 1 \quad \text{e} \quad \frac{g(x+h)}{g(x)} = a^h$$

dependem apenas de h, mas não de x. Mostraremos agora que vale a recíproca.

Teorema: (Caracterização das funções de tipo exponencial.) *Seja* $g: \mathbb{R} \to \mathbb{R}^+$ *uma função monótona injetiva (isto é, crescente ou decrescente) tal que, para* $x, h \in \mathbb{R}$ *quaisquer, o acréscimo relativo* $[g(x+h) - g(x)]/g(x)$ *dependa apenas de* h*, mas não de* x*. Então, se* $b = g(0)$ *e* $a = g(1)/g(0)$*, tem-se* $g(x) = ba^x$ *para todo* $x \in \mathbb{R}$.

Demonstração: Como vimos acima, a hipótese feita equivale a supor que $\varphi(h) = g(x+h)/g(x)$ independe de x. Substituindo, se necessário, $g(x)$ por $f(x) = g(x)/b$, onde $b = g(0)$, f continua monótona injetiva, com $f(x+h)/f(x)$ independente de x e, agora, com $f(0) = 1$. Então, pondo $x = 0$ na relação $\varphi(h) = f(x+h)/f(x)$, obtemos $\varphi(h) = f(h)$ para todo $h \in \mathbb{R}$. Vemos assim que a função monótona injetiva f cumpre $f(x+h) = f(x) \cdot f(h)$, ou seja $f(x+y) = f(x) \cdot f(y)$ para quaisquer $x, y \in \mathbb{R}$. Segue-se então do teorema anterior que $f(x) = a^x$, logo $g(x) = bf(x) = ba^x$, como queríamos demonstrar.

5. Funções Exponenciais e Progressões

Seja $f: \mathbb{R} \to \mathbb{R}$, $f(x) = ba^x$, uma função de tipo exponencial. Se $x_1, x_2, \ldots, x_n, \ldots$ é uma progressão aritmética de razão h, isto é, $x_{n+1} = x_n + h$, então os valores

$$f(x_1) = ba^{x_1}, f(x_2) = ba^{x_2}, \ldots, f(x_n) = ba^{x_n}, \ldots,$$

formam uma progressão geométrica de razão a^h pois

$$f(x_{n+1}) = ba^{x_{n+1}} = ba^{x_n+h} = (ba^{x_n}) \cdot a^h.$$

Como o $(n+1)$-ésimo termo da progressão aritmética dada é $x_{n+1} = x_1 + nh$, segue-se que $f(x_{n+1}) = f(x_1) \cdot A^n$, onde $A = a^h$. Em particular, se $x_1 = 0$ então $f(x_1) = b$ logo $f(x_{n+1}) = b \cdot A^n$.

Esta simples observação é usada na prática para "discretizar" a análise das situações, como aquelas da seção 1, em que se tem crescimento ou decrescimento exponencial.

Por exemplo, se um capital inicial c_0 é aplicado a juros fixos então, depois de decorrido um tempo t, o capital existente é dado por $c(t) = c_0 \cdot a^t$. Se tirarmos extratos da conta nos tem-

pos $0, h, 2h, 3h, \ldots$ teremos $c(0) = c_0$, $c(h) = c_0 A$, $c(2h) = c_0 \cdot A^2$, $c(3h) = c_0 \cdot A^3, \ldots$ onde $A = a^h$. Portanto, a evolução do saldo, quando calculado em intervalos de h unidades de tempo, é dada pela progressão geométrica:

$$c_0, c_0 \cdot A, c_0 \cdot A^2, c_0 \cdot A^3, \ldots .$$

(Vide "Progressões e Matemática Financeira", Coleção do Professor de Matemática, SBM)

Esta propriedade é característica das funções de tipo exponencial, conforme o

Teorema: *Seja* $f: \mathbb{R} \to \mathbb{R}^+$ *uma função monótona injetiva (isto é, crescente ou decrescente) que transforma toda progressão aritmética* $x_1, x_2, \ldots, x_n, \ldots$ *numa progressão geométrica* $y_1, y_2, \ldots, y_n, \ldots, y_n = f(x_n)$. *Se pusermos* $b = f(0)$ *e* $a = f(1)/f(0)$ *teremos* $f(x) = ba^x$ *para todo* $x \in \mathbb{R}$.

Demonstração: Seja $b = f(0)$. A função $g: \mathbb{R} \to \mathbb{R}^+$, definida por $g(x) = f(x)/b$, é monótona injetiva, continua transformando progressões aritméticas em progressões geométricas e agora tem-se $g(0) = 1$. Dado $x \in \mathbb{R}$ qualquer, a seqüência $x, 0, -x$ é uma progressão aritmética, logo $g(x), 1, g(-x)$ é uma progressão geométrica de razão $g(-x)$. Segue-se $g(-x) = 1/g(x)$. Sejam agora $n \in \mathbb{N}$ e $x \in \mathbb{R}$. A seqüência $0, x, 2x, \ldots, nx$ é uma progressão aritmética, logo $1, g(x), g(2x), \ldots, g(nx)$ é uma progressão geométrica, cuja razão evidentemente é $g(x)$. Então seu $(n+1)$-ésimo termo é $g(nx) = g(x)^n$. Se $-n$ é um inteiro negativo então $g(-nx) = 1/g(nx) = 1/g(x)^n = g(x)^{-n}$. Portanto, vale $g(nx) = g(x)^n$ para quaisquer $n \in \mathbb{Z}$ e $x \in \mathbb{R}$. Segue-se do Teorema de Caracterização acima que, pondo $a = g(1) = f(1)/f(0)$, tem-se $g(x) = a^x$, ou seja, $f(x) = ba^x$, para todo $x \in \mathbb{R}$.

6. Função Inversa

Diz-se que a função $g: Y \to X$ é a *inversa* da função $f: X \to Y$ quando

se tem $g(f(x)) = x$ e $f(g(y)) = y$ para quaisquer $x \in X$ e $y \in Y$. Evidentemente, g é inversa de f se, e somente se, f é inversa de g.

Quando g é a inversa de f, tem-se $g(y) = x$ se, e somente se, $f(x) = y$.

Se $g(f(x)) = x$ para todo $x \in X$ então a função f é injetiva, pois

$$f(x_1) = f(x_2) \Rightarrow g(f(x_1)) = g(f(x_2)) \Rightarrow x_1 = x_2\,.$$

Por sua vez, a igualdade $f(g(y)) = y$, valendo para todo $y \in Y$, implica que f é sobrejetiva pois, dado $y \in Y$ arbitrário, tomamos $x = g(y) \in X$ e temos $f(x) = y$.

Portanto, se a função $f\colon X \to Y$ possui inversa então f é injetiva e sobrejetiva, ou seja, é uma correspondência biunívoca entre X e Y.

Reciprocamente, se $f\colon X \to Y$ é uma correspondência biunívoca entre X e Y então f possui uma inversa $g\colon Y \to X$. Para definir g, notamos que, sendo f sobrejetiva, para todo $y \in Y$ existe algum $x \in X$ tal que $f(x) = y$. Além disso, como f é injetiva, este x é único. Pomos então $g(y) = x$. Assim, $g\colon Y \to X$ é a função que associa a cada $y \in Y$ o único $x \in X$ tal que $f(x) = y$. É imediato que $g(f(x)) = x$ e $f(g(y)) = y$ para $x \in X$ e $y \in Y$ quaisquer.

Exemplo. Lembremos que $[0, +\infty) = \{x \in \mathbb{R}; x \geqslant 0\}$. Sejam $f\colon \mathbb{R} \to [0, +\infty)$ e $g\colon [0, +\infty) \to \mathbb{R}$ definidas por $f(x) = x^2$ e $g(y) = \sqrt{y}$. Tem-se $f(g(y)) = y$ para todo $y \geqslant 0$ mas $g(f(x))$ só é igual a x quando $x \geqslant 0$. Se $x \in \mathbb{R}$ for negativo então $g(f(x)) = -x$. Portanto g não é inversa de f. Na realidade, nenhuma função $\varphi\colon [0, +\infty) \to \mathbb{R}$ pode ser inversa de f porque f não é injetiva. Note, porém, que se considerarmos a restrição de f a $[0, +\infty)$, isto é, a função $F\colon [0, +\infty) \to [0, +\infty)$, dada por $F(x) = x^2$, então F é uma correspondência biunívoca, e sua inversa é a função $G\colon [0, +\infty) \to [0, +\infty)$, dada por $G(y) = \sqrt{y}$, pois

$$G(F(x)) = G(x^2) = \sqrt{x^2} = x$$

e

$$F(G(y)) = F(\sqrt{y}) = (\sqrt{y})^2 = y$$

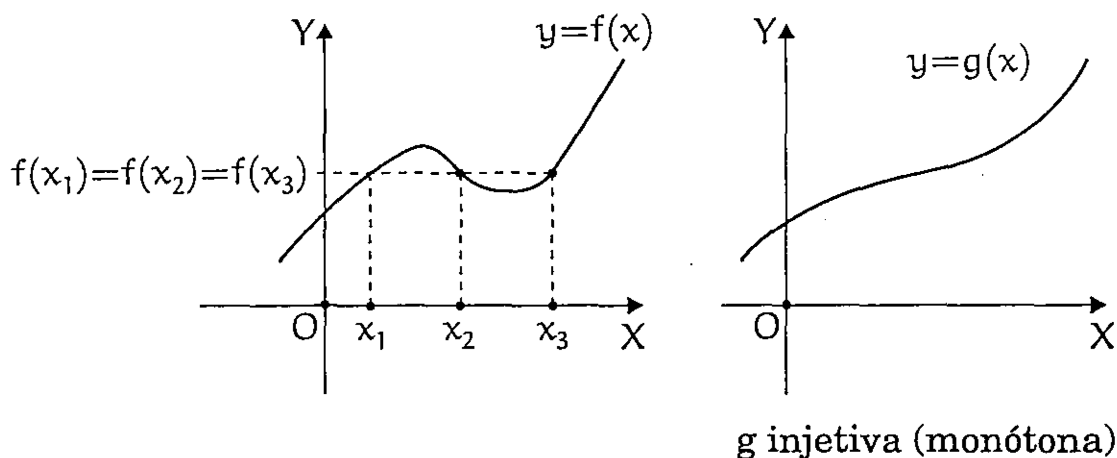
para quaisquer $x \geqslant 0$ e $y \geqslant 0$.

Mais geralmente, para todo $n \in \mathbb{N}$, a função $x \mapsto x^n$ é uma correspondência biunívoca de $[0, +\infty)$ sobre si mesmo, cuja inversa é $y \mapsto \sqrt[n]{y}$.

Se n é ímpar, então $x \mapsto x^n$ é uma correspondência biunívoca de $\mathbb{R}$ sobre si mesmo, cuja inversa $G: \mathbb{R} \to \mathbb{R}$ é dada por $G(y) = \sqrt[n]{y}$.

Quando $g: Y \to X$ é a função inversa de $f: X \to Y$, escreve-se $g = f^{-1}$.

Prova-se que uma função contínua $f: I \to \mathbb{R}$, definida num intervalo $I \subset \mathbb{R}$, só pode ser injetiva se for monótona (crescente ou decrescente).



g injetiva (monótona)

Figura 57

Portanto, a fim de que uma função contínua $f: I \to J$ (I, J intervalos) possua uma inversa, é necessário que f seja crescente, ou decrescente, além de sobrejetiva.

A inversa de uma função crescente é crescente e a inversa de uma função decrescente é decrescente.

Antes de falar sobre o gráfico da função inversa, revejamos a noção de simetria em relação a uma reta.

Dois pontos P, Q no plano dizem-se *simétricos* em relação a uma reta r nesse plano quando r é a mediatriz do segmento PQ. Duas figuras dizem-se simétricas em relação à reta r quando cada ponto de uma delas é o simétrico de um ponto da outra em relação a essa reta.

Chama-se *diagonal* do plano $\mathbb{R}^2$ a reta Δ formada pelos pontos (x, x) que têm abcissa e ordenada iguais.

O simétrico do ponto $P = (x, y) \in \mathbb{R}^2$ em relação à diagonal Δ é o ponto $Q = (y, x)$. Com efeito, o segmento PQ é uma diagonal do quadrado cujos vértices são (x, y), (x, x), (y, x) e (y, y), enquanto Δ é o prolongamento da outra diagonal.

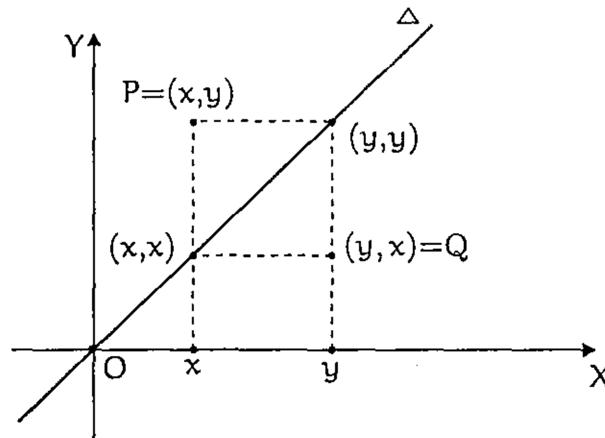


Figura 58

Se X, Y são conjuntos de números reais e $f^{-1}: Y \to X$ é a inversa da função $f: X \to Y$ então o gráfico G' da função f^{-1} é o simétrico do gráfico G da função f em relação à diagonal $\Delta \subset \mathbb{R}^2$.

Com efeito, temos

$$(x, y) \in G \quad \Leftrightarrow \quad y = f(x) \quad \Leftrightarrow \quad x = f^{-1}(y) \quad \Leftrightarrow \quad (y, x) \in G'.$$

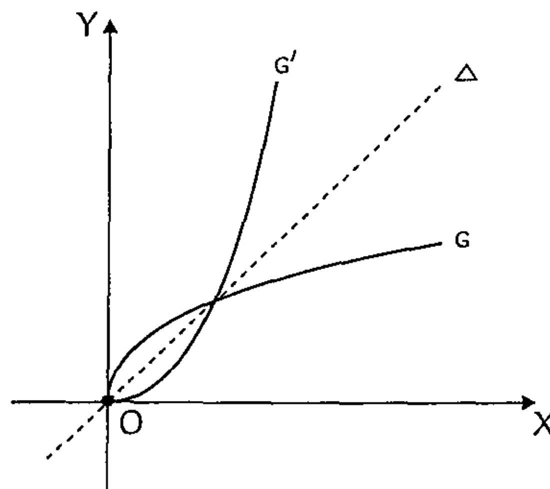


Figura 59

Se, numa folha de papel translúcido, traçarmos o gráfico de uma função f então, girando a folha num ângulo de 180° em torno da diagonal Δ obtemos o gráfico de f^{-1}.

Observação. Se $f: X \to Y$ é sobrejetiva e $g: Y \to X$ é tal que $g(f(x)) = x$ para todo $x \in X$ então tem-se necessariamente $f(g(y)) = y$ para todo $y \in Y$ e $g = f^{-1}$ é a inversa de x. Com efeito, dado qualquer $y \in Y$ existe $x \in X$ tal que $f(x) = y$, logo

$$f(g(y)) = f(g(f(x))) = f(x) = y.$$

7. Funções Logarítmicas

Vimos na seção 5 que, para todo número real positivo $a \neq 1$, a função exponencial $f: \mathbb{R} \to \mathbb{R}^+$, $f(x) = a^x$, é uma correspondência biunívoca entre $\mathbb{R}$ e $\mathbb{R}^+$, crescente se $a > 1$, decrescente se $0 < a < 1$, com a propriedade adicional

$$f(x + y) = f(x) \cdot f(y).$$

Segue-se que f possui uma função inversa.

A inversa da função exponencial de base a é a função

$$\log_a: \mathbb{R}^+ \to \mathbb{R},$$

que associa a cada número real positivo x o número real $\log_a x$, chamado o *logaritmo* de x na base a. Por definição de função inversa, tem-se

$$a^{\log_a x} = x \quad \text{e} \quad \log(a^x) = x.$$

Assim, $\log_a x$ é o expoente ao qual se deve elevar a base a para obter o número x. Ou seja,

$$y = \log_a x \quad \Leftrightarrow \quad a^y = x.$$

Segue-se imediatamente da relação $a^u \cdot a^v = a^{u+v}$ que

$$\log_a(xy) = \log_a x + \log_a y$$

para x e y positivos quaisquer. Com efeito, se $u = \log_a x$ e $v = \log_a y$

então $a^u = x$ e $a^v = y$, logo

$$xy = a^u \cdot a^v = a^{u+v},$$

ou seja

$$\log_a(xy) = u + v = \log_a x + \log_a y.$$

Esta propriedade de transformar produtos em somas foi a motivação original para a introdução dos logaritmos, no início do século 17, e de sua popularidade, até bem recentemente, como um eficiente instrumento de cálculo.

O uso generalizado das calculadoras, cada vez mais desenvolvidas, fez com que essa utilidade inicial dos logaritmos perdesse o sentido. Entretanto, a função logaritmo continua extremamente importante na Matemática e em suas aplicações.

Essa importância é permanente; jamais desaparecerá porque, sendo a inversa da função exponencial (portanto equivalente a ela), a função logaritmo está ligada a um grande número de fenômenos e situações naturais, onde se tem uma grandeza cuja taxa de variação é proporcional à quantidade da mesma existente no instante dado. (Vide RPM 18, pág. 24 e o livro "Logaritmos", já citado.)

A função $\log_a\colon \mathbb{R}^+ \to \mathbb{R}$ é crescente quando $a > 1$ e decrescente quando $0 < a < 1$. Como $a^0 = 1$, tem-se $\log_a 1 = 0$. É importante ressaltar que somente números positivos possuem logaritmo real, pois a função $x \mapsto a^x$ somente assume valores positivos. (Para uma discussão sobre logaritmos de números negativos, ver "Meu Professor de Matemática", página 180.)

As funções logarítmicas mais utilizadas são aquelas de base $a > 1$, especialmente as de base 10 (logaritmos *decimais*), base 2 (logaritmos *binários*) e base e (logaritmos *naturais*, às vezes impropriamente chamados *neperianos*). Estes últimos são os mais adequados cientificamente, e voltaremos a eles logo mais.

Como $\log_a x$ é uma função crescente de x quando $a > 1$, e como $\log_a 1 = 0$, segue-se que, para $a > 1$, os números compreendidos entre 0 e 1 têm logaritmo negativo e os maiores do que 1 têm lo-

garitmo positivo. Ao contrário, se $0 < a < 1$ então $\log_a x$ é positivo quando $0 < x < 1$ e negativo quando $x > 1$. A figura mostra os gráficos das funções $f(x) = \log_2 x$ e $g(x) = \log_{1/2} x$.

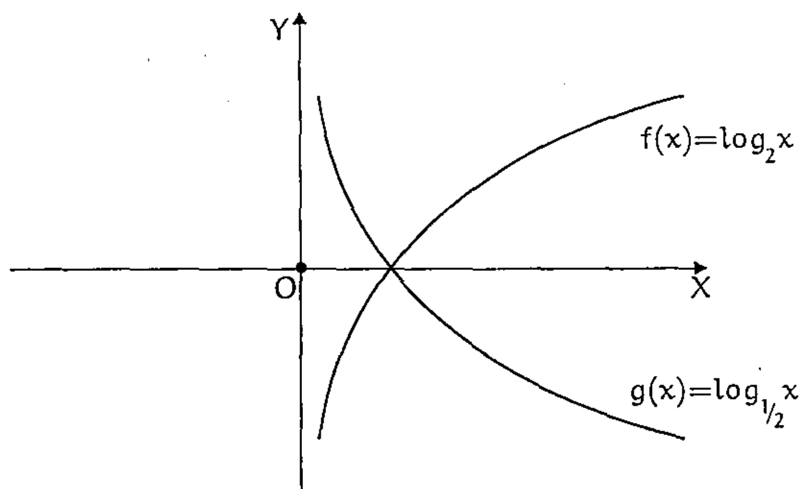


Figura 60

Se tivéssemos traçado os gráficos das funções $y = \log_a x$ e $y = \log_b x$, com $a > 1$ e $0 < b < 1$ quaisquer, as figuras obtidas teriam o mesmo aspecto acima. Mais precisamente, existiriam constantes positivas c, d tais que $\log_a x = c \cdot \log_2 x$ e $\log_b x = d \cdot \log_{1/2} x$ para todo $x > 0$.

Com efeito se $u = \log_a x$ e $v = \log_2 x$ então $a^u = x$ e $2^v = x$. Portanto, se escrevermos $c = \log_a 2$ teremos $a^c = 2$, logo

$$x = a^u = 2^v = (a^c)^v = a^{cv}$$

portanto $u = cv$, isto é, $\log_a x = c \cdot \log_2 x$ para todo $x > 0$, onde a constante c é igual a $\log_a 2$. A igualdade

$$\log_a x = \log_a b \cdot \log_b x$$

é válida em geral (mesmo raciocínio) e se chama a *fórmula de mudança de base* para logaritmos. Quando a e b são ambos maiores ou ambos menores do que 1 então $\log_a b > 0$. Se um dos números a, b é maior e o outro é menor do que 1 então $\log_a b < 0$. A fórmula acima diz que duas funções logarítmicas quaisquer diferem por um fator constante.

Como $\log_a : \mathbb{R}^+ \to \mathbb{R}$ é uma correspondência biunívoca, portanto sobrejetiva, segue-se que $y = \log_a x$ é uma função ilimitada, tanto superiormente quanto inferiormente. Mais precisamente, tem-se, para $a > 1$:

$$\lim_{x \to +\infty} \log_a x = +\infty \quad \text{e} \quad \lim_{x \to 0} \log_a x = -\infty.$$

A primeira destas igualdades significa que se pode dar a $\log_a x$ um valor tão grande quanto se queira, desde que x seja tomado suficientemente grande. A segunda quer dizer que, dado arbitrariamente $A > 0$, tem-se $\log_a x < -A$ desde que x seja um número positivo suficientemente pequeno.

Ao contrário da função exponencial, que cresce rapidamente, $\log_a x$ tende a $+\infty$ muito lentamente quando $x \to +\infty$. Com efeito, dado um número $M > 0$, tem-se $\log_a x > M \Leftrightarrow x > a^M$. Assim, por exemplo, se quisermos que $\log_{10} x$ seja maior do que mil, será preciso tomar um número x cuja expressão decimal tenha pelo menos mil e um algarismos.

Esse crescimento lento do logaritmo, que contrasta com o crescimento rápido da exponencial, é bem ilustrado pelos gráficos das funções $y = a^x$ e $y = \log_a x$, que, como sabemos, são simétricos em relação à diagonal de $\mathbb{R}^2$.

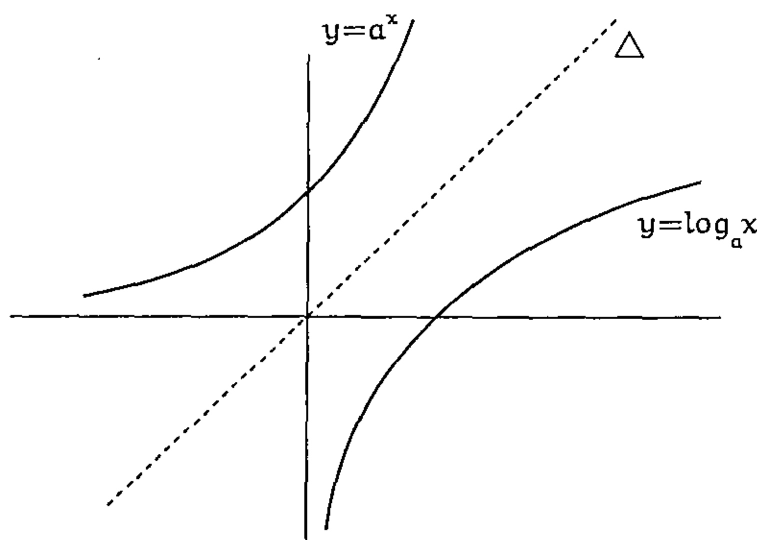


Figura 61

8. Caracterização das Funções Logarítmicas

Provaremos a seguir que, entre as funções monótonas injetivas $\mathbb{R}^+ \to \mathbb{R}$, somente as funções logarítmicas têm a propriedade de transformar produtos em somas. Antes lembremos que se $f\colon \mathbb{R}^+ \to \mathbb{R}$ é tal que $f(a^x) = x$ para todo $x \in \mathbb{R}$ então $f(y) = \log_a y$ para todo $y \in \mathbb{R}^+$, de acordo com a Observação no final da seção 6, pois $x \mapsto a^x$ é uma função sobrejetiva de $\mathbb{R}$ em $\mathbb{R}^+$. (Estamos supondo $a > 0$ diferente de 1.)

Teorema: (Caracterização das funções logarítmicas.) *Seja* $f\colon \mathbb{R}^+ \to \mathbb{R}$ *uma função monótona injetiva (isto é, crescente ou decrescente) tal que* $f(xy) = f(x) + f(y)$ *para quaisquer* $x, y \in \mathbb{R}^+$. *Então existe* $a > 0$ *tal que* $f(x) = \log_a x$ *para todo* $x \in \mathbb{R}^+$.

Demonstração: Para fixar as idéias, admitamos f crescente. O outro caso é tratado igualmente. Temos $f(1) = f(1 \cdot 1) = f(1) + f(1)$, logo $f(1) = 0$. Provemos o teorema inicialmente supondo que exista $a \in \mathbb{R}^+$ tal que $f(a) = 1$. Depois mostraremos que isto sempre acontece, logo não é uma hipótese adicional. Como f é crescente e $f(a) = 1 > 0 = f(1)$, tem-se $a > 1$. Para todo $m \in \mathbb{N}$ vale

$$\begin{aligned} f(a^m) &= f(a \cdot a \cdot \ldots \cdot a) \\ &= f(a) + f(a) + \cdots + f(a) \\ &= 1 + 1 + \cdots + 1 = m, \end{aligned}$$

$$\begin{aligned} 0 = f(1) &= f(a^m \cdot a^{-m}) \\ &= f(a^m) + f(a^{-m}) = m + f(a^{-m}), \end{aligned}$$

logo $f(a^{-m}) = -m$. Se $r = m/n$ com $m \in \mathbb{Z}$ e $n \in \mathbb{N}$ então $rn = m$, portanto

$$m = f(a^m) = f(a^{rn}) = f((a^r)^n) = n \cdot f(a^r)$$

e daí $f(a^r) = \frac{m}{n} = r$. Se $x \in \mathbb{R}$ é irracional então, para r, s racionais tem-se:

$$r < x < s \Rightarrow a^r < a^x < a^s \Rightarrow f(a^r) < f(a^x) < f(a^s) \Rightarrow r < f(a^x) < s.$$

Assim todo número racional r, menor do que x, é também menor do que $f(a^x)$ e todo número racional s maior do que x é também maior do que $f(a^x)$. Segue-se que $f(a^x) = x$ para todo $x \in \mathbb{R}$. Portanto $f(y) = \log_a y$ para todo $y > 0$.

Consideremos agora o caso geral, em que se tem uma função crescente $g: \mathbb{R}^+ \to \mathbb{R}$, tal que

$$g(xy) = g(x) + g(y),$$

sem mais nenhuma hipótese. Então $g(1) = 0$ e, como $1 < 2$, devemos ter $g(2) = b > 0$. A nova função $f: \mathbb{R}^+ \to \mathbb{R}$, definida por $f(x) = g(x)/b$, é crescente, transforma somas em produtos e cumpre $f(2) = 1$. Logo, pela primeira parte da demonstração, tem-se $f(x) = \log_2 x$ para todo $x > 0$. Isto significa que, para todo $x > 0$, vale

$$x = 2^{f(x)} = 2^{g(x)/b} = (2^{1/b})^{g(x)} = a^{g(x)},$$

com $a = 2^{1/b}$. Tomando $\log_a$ de ambos os membros da igualdade $a^{g(x)} = x$ vem, finalmente: $g(x) = \log_a x$.

9. Logaritmos Naturais

Nesta seção, mostraremos como os logaritmos naturais podem ser apresentados de forma geométrica, usando para isso o Teorema de Caracterização demonstrado na seção anterior.

Começamos pelo estudo de uma transformação geométrica bastante simples, que se revela útil para os nossos propósitos.

Para cada número real $k > 0$, definimos a transformação (= função) $T = T_k: \mathbb{R}^2 \to \mathbb{R}^2$, que associa a cada ponto $(x, y) \in \mathbb{R}^2$ o ponto $T(x, y) = (kx, y/k)$, obtido de (x, y) multiplicando a abcissa por k e dividindo a ordenada pelo mesmo k.

Um retângulo X de lados paralelos aos eixos, com base medindo b e altura medindo a, é transformado por T num retângulo $X' = T(X)$, ainda com lados paralelos aos eixos, porém com base kb e altura a/k. Portanto X e seu transformado $X' = T(X)$ têm áreas iguais. Mais geralmente, T transforma toda figura F do plano numa figura $F' = T(F)$, cujas dimensões em relação a F são altera-

das pelo fator k na horizontal e $1/k$ na vertical. Logo F e F' têm a mesma área.

O leitor interessado numa análise mais detida do fato de que F e F' têm a mesma área observará que todo polígono retangular contido em F é transformado por T num polígono retangular de mesma área contido em F' enquanto T^{-1} faz o mesmo com os polígonos retangulares contido em F'. [Vide "Medida e Forma em Geometria", especialmente as pags. 22 e 49.]

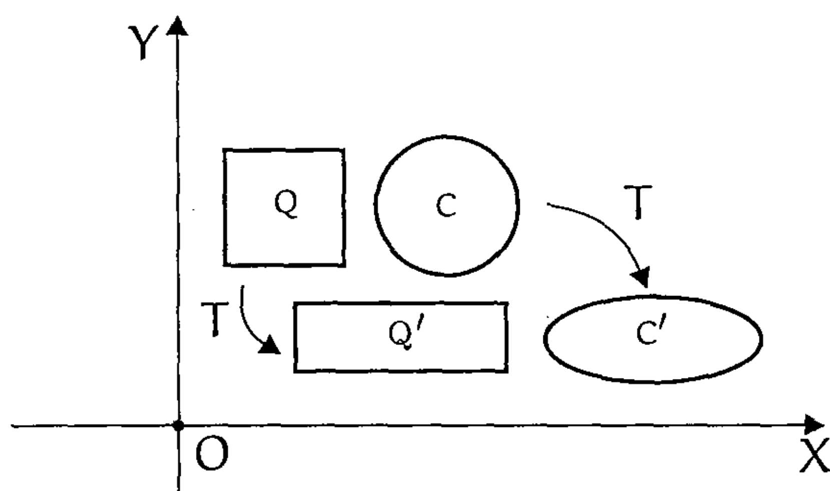


Figura 62: Um quadrado, um círculo e suas imagens por $T(x, y) = (2x, y/2)$

Interessa-nos em particular o efeito da transformação T nas faixas de hipérbole.

Seja

$$H = \{(x, 1/x); x > 0\}$$

o ramo positivo da hipérbole equilátera $xy = 1$; H é o gráfico da função $h\colon \mathbb{R}^+ \to \mathbb{R}$, $h(x) = 1/x$.

Dados $a, b \in \mathbb{R}^+$, o conjunto H_a^b dos pontos (x, y) do plano tais que x está entre a e b e $0 \leqslant y \leqslant 1/x$ chama-se uma *faixa de hipérbole*. H_a^b é o conjunto do plano limitado lateralmente pelas verticais $x = a$, $x = b$, ao sul pelo eixo das abcissas e ao norte pela hipérbole H.

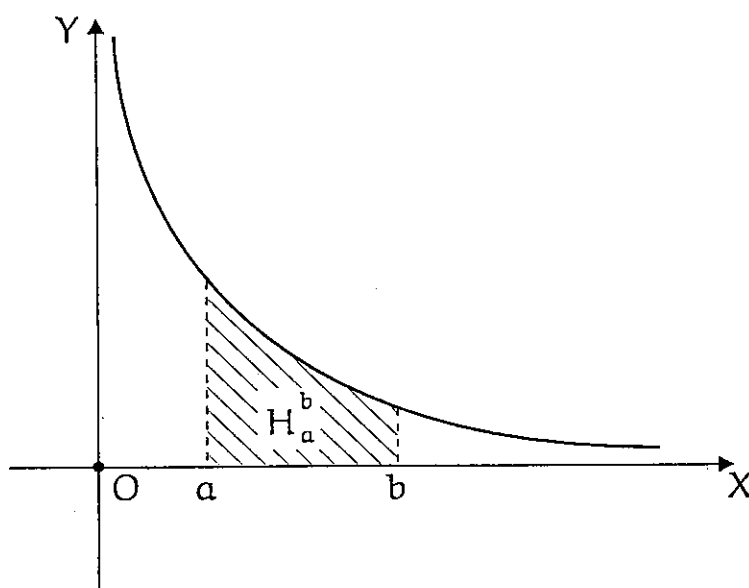


Figura 63

A transformação $T = T_k: \mathbb{R}^2 \to \mathbb{R}^2$ leva a faixa H_b^b na faixa H_{ak}^{bk}.

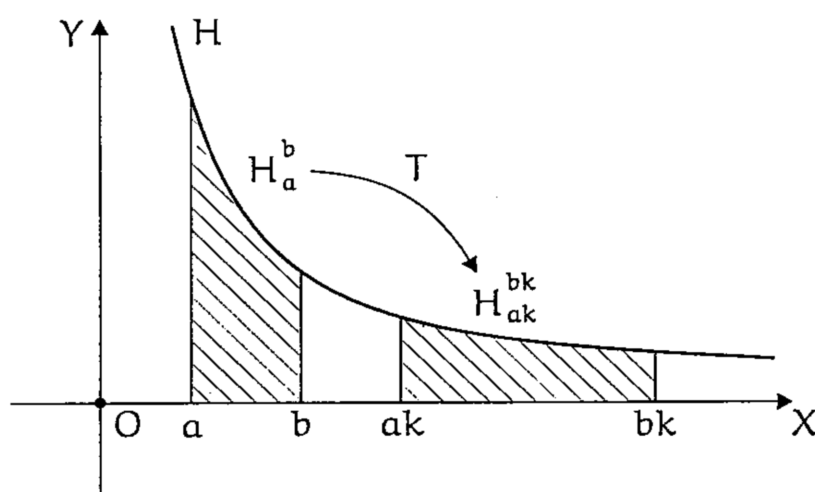


Figura 64

Como T preserva áreas, segue-se que, para todo $k > 0$, as faixas H_a^b e H_{ak}^{bk} têm a mesma área.

Normalmente, a área de uma figura não é um número negativo. Mas às vezes é conveniente usar "áreas orientadas", ou seja, providas de sinal $+$ ou $-$. É o que faremos agora.

Convencionaremos que a área da faixa de hipérbole H_a^b será positiva quando $a < b$, negativa quando $b < a$ e zero quando

$a = b$.

Para deixar mais clara esta convenção, escreveremos

$$\text{ÁREA } H_a^b,$$

com letras maiúsculas, para indicar a área orientada (provida de sinal). A área usual, com valores $\geqslant 0$, será escrita como área H_a^b. Assim, temos

$$\text{ÁREA } H_a^b = \text{área } H_a^b > 0 \text{ se } a < b;$$

$$\text{ÁREA } H_a^b = \text{-área } H_a^b < 0 \text{ se } b < a;$$

$$\text{ÁREA } H_a^a = 0.$$

É óbvio que, quando $a < b < c$, tem-se

$$\text{área } H_a^b + \text{ área } H_b^c = \text{ área } H_a^c.$$

Uma conseqüência da adoção de áreas orientadas é que se tem

$$\text{ÁREA } H_a^b = - \text{ ÁREA } H_b^a.$$

Daí segue que vale a igualdade

$$\text{ÁREA } H_a^b + \text{ ÁREA } H_b^c = \text{ ÁREA } H_a^c$$

em qualquer dois seis casos $a \leqslant b \leqslant c$, $a \leqslant c \leqslant b$, $b \leqslant a \leqslant c$, $b \leqslant c \leqslant a$, $c \leqslant a \leqslant b$ e $c \leqslant b \leqslant a$. A igualdade acima é fácil de provar. Basta ter a paciência de considerar separadamente cada uma destas seis possibilidades.

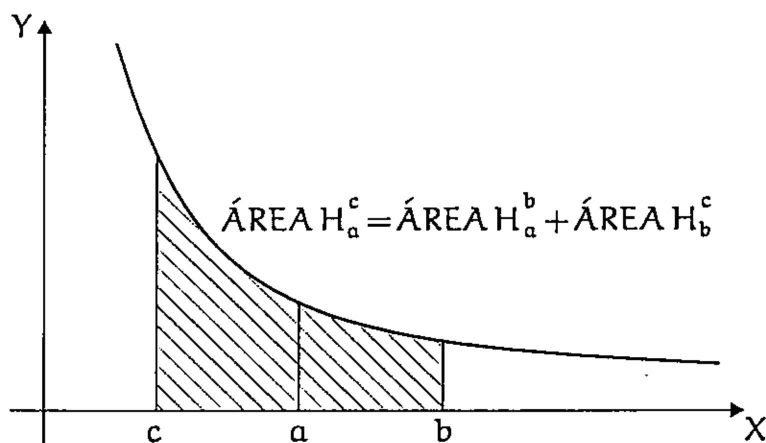


Figura 65

Definamos uma função $f: \mathbb{R}^+ \to \mathbb{R}$ pondo, para cada número real $x > 0$.

$$f(x) = \text{ÁREA } H_1^x$$

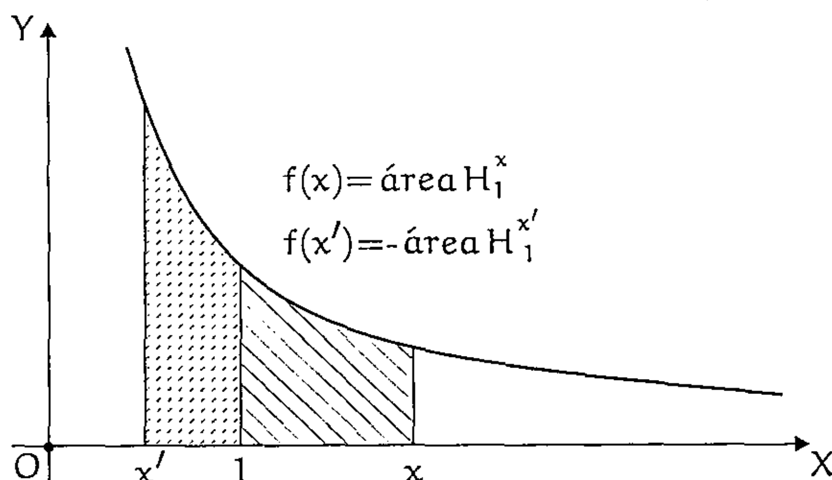


Figura 66

$\ell n\, x =$ área da região hachurada
$\ell n\, x' =$ - área da região pontilhada

Resultam imediatamente da definição as seguintes propriedades:

$$f(x) > 0 \quad \Leftrightarrow \quad x > 1;$$
$$f(x) < 0 \quad \Leftrightarrow \quad 0 < x < 1;$$
$$f(1) = 0;$$
f é crescente.

Além disso, observamos que, para $x, y \in \mathbb{R}^+$ quaisquer:

$$f(xy) = \text{ÁREA } H_1^{xy} = \text{ÁREA } H_1^x + \text{ÁREA } H_x^{xy}.$$

Mas, como vimos acima, $\text{ÁREA } H_x^{xy} = \text{ÁREA } H_1^y$. Logo $f(xy) = \text{ÁREA } H_1^x + \text{ÁREA } H_1^y$, ou seja:

$$f(xy) = f(x) + f(y).$$

Pelo Teorema de Caracterização das funções logarítmicas, existe um número real positivo, que chamaremos de e, tal que

$$f(x) = \log_e x \quad \text{para todo} \quad x \in \mathbb{R}^+.$$

Escreveremos $\ell n\, x$ em vez de $\log_e x$ e chamaremos o número $\ell n\, x$ de *logaritmo natural* de x

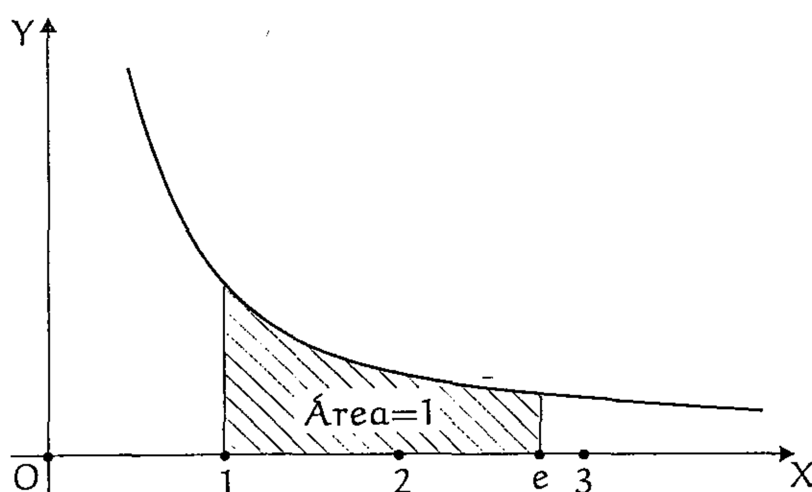


Figura 67

O número e, base dos logaritmos naturais, é caracterizado pelo fato de que seu logaritmo natural é igual a 1, ou seja ÁREA $H_1^e = 1$.

O número e é irracional. Um valor aproximado dessa importante constante é $e = 2,718281828459$.

Os logaritmos naturais, de base e, são os mais importantes nas aplicações, especialmente aquelas que envolvem o uso do Cálculo Infinitesimal.

Alguns autores chamam o logaritmo natural de "logaritmo neperiano", em homenagem a John Napier, autor da primeira tábua de logaritmos, em 1614. Entretanto, tal denominação não é inteiramente apropriada, pois o logaritmo originalmente definido por Napier não coincide com o logaritmo natural.

Usualmente, o número e é apresentado como o limite da expressão $\left(1 + \frac{1}{n}\right)^n$ quando n tende ao infinito. Noutras palavras, costuma-se introduzir e como o número real cujos valores aproximados por falta são os números racionais da forma $\left(1 + \frac{1}{n}\right)^n$,



$n \in \mathbb{N}$. Essas aproximações são tanto melhores quanto maior for o número n. Mostraremos agora que o número e, que acabamos de caracterizar pela propriedade ÁREA $H_1^e = 1$, é mesmo o valor daquele limite.

O argumento que usaremos para dar essa prova se baseia na figura abaixo, copiada da capa do livro "Logaritmos", já citado antes.

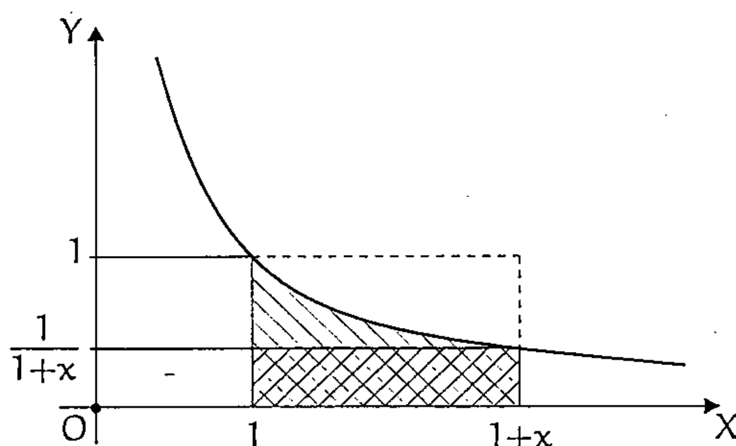


Figura 68

Nela temos um retângulo menor, cuja base mede x e cuja altura mede $\frac{1}{1+x}$, contido na faixa H_1^{1+x} e esta faixa, por sua vez, contida no retângulo maior, com a mesma base de medida x e altura igual a 1. Comparando as áreas dessas três figuras, podemos escrever, para todo $x > 0$:

$$\frac{x}{1+x} < \ell n(1+x) < x.$$

Dividindo por x:

$$\frac{1}{1+x} < \frac{\ell n(1+x)}{x} < 1.$$

Tomando $x = \frac{1}{n}$:

$$\frac{n}{n+1} < \ell n\left(1+\frac{1}{n}\right)^n < 1,$$

portanto:

$$e^{\frac{n}{n+1}} < \left(1+\frac{1}{n}\right)^n < e,$$

para todo $n \in \mathbb{N}$. Quando n cresce indefinidamente, $\frac{n}{n+1}$ se aproxima de 1, logo $e^{\frac{n}{n+1}}$ tende a e. Segue-se então destas últimas desigualdades que

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e.$$

Este argumento ilustra bem claramente a vantagem que advém de se interpretar o logaritmo natural geometricamente: a noção de área é visualmente intuitiva, permititndo que se obtenham desigualdades como a que foi usada aqui.

A igualdade $e = \lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^n$ foi obtida a partir da desigualdade

$$(*) \qquad \frac{1}{1+x} < \frac{\ell n(1+x)}{x} < 1,$$

válida para todo $x > 0$. Se considerarmos $-1 < x < 0$, teremos $-x > 0$ e $1+x > 0$. Portanto é válido ainda falar de $\ell n(1+x)$. Observamos que o retângulo cuja base mede $-x$ e cuja altura mede 1 está contido na faixa H_{1+x}^{1} e esta, por sua vez, está contida no retângulo de mesma base e altura $1/(1+x)$. Comparando as áreas destas figuras, vem

$$-x < -\ell n(1+x) < -\frac{x}{1+x}.$$

Dividindo os 3 membros pelo número positivo $-x$ obtemos

$$(**) \qquad 1 < \frac{\ell n(1+x)}{x} < \frac{1}{1+x}.$$

As desigualdades (*) e (**) nos dão

$$\frac{1}{1+x} < \ell n(1+x)^{\frac{1}{x}} < 1 \quad \text{ou} \quad 1 < \ell n(1+x)^{\frac{1}{x}} < \frac{1}{1+x},$$

ou seja

$$e^{\frac{1}{1+x}} < (1+x)^{\frac{1}{x}} < e \quad \text{ou} \quad e < (1+x)^{\frac{1}{x}} < e^{\frac{1}{1+x}},$$

conforme seja $x > 0$ ou $-1 < x < 0$. Em qualquer hipótese, daí se segue que

(***) $$\lim_{x \to 0} (1 + x)^{\frac{1}{x}} = e.$$

Isto significa que é possível tornar o valor da expressão $(1+x)^{\frac{1}{x}}$ tão próximo de e quanto se deseje, desde que se torne o número não-nulo x suficientemente pequeno em valor absoluto. (O próprio x pode ser > 0 ou < 0.)

A igualdade (***) se exprime dizendo que $(1+x)^{1/x}$ tende a e quando x tende a zero.

Tomando, por exemplo, $x = \frac{\alpha}{n}$, vemos que $\frac{1}{x} = \frac{n}{\alpha}$ e que $x \to 0$ se, e somente se $n \to \infty$. Logo (***) nos dá

$$\lim_{n \to \infty} \left(1 + \frac{\alpha}{n}\right)^n = \lim_{n \to \infty} \left[\left(1 + \frac{\alpha}{n}\right)^{\frac{n}{\alpha}}\right]^\alpha = \lim_{x \to 0} \left[(1+x)^{\frac{1}{x}}\right]^\alpha = e^\alpha.$$

Como caso particular da igualdade

$$e^\alpha = \lim_{n \to \infty} \left(1 + \frac{\alpha}{n}\right)^n,$$

válida para todo $\alpha \in \mathbb{R}$, obtemos

$$\frac{1}{e} = \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^n.$$

10. A Função Exponencial de Base e

O número e, base dos logaritmos naturais, foi definido na seção anterior como o único número real positivo tal que a área da faixa de hipérbole H_1^e é igual a 1. Em seguida, mostramos que esse número é também o limite de $\left(1 + \frac{1}{n}\right)^n$ quando n tende ao infinito. Nesta seção, daremos exemplo de uma situação da vida real que leva à consideração do limite acima.

Por sua vez, a função exponencial $x \mapsto e^x$, de base e, pode ser definida por meio do limite $e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$ ou então, geometricamente, pelo fato de que $y = e^x$ é o único número real positivo tal

que a área da faixa de hipérbole H_1^y é igual a x. Mostraremos que as funções de tipo exponencial, $f(x) = b\, e^{\alpha x}$, com base e, surgem em questões naturais e calcularemos a taxa de variação instantânea dessas funções.

Um investidor aplica um capital c_0 a uma taxa de k por cento ao ano. Se escrevermos, por simplicidade, $\alpha = k/100$, por cada real aplicado o investidor receberá, no final de um ano, $1 + \alpha$ reais, de modo que o total a ser resgatado será $c_0(1 + \alpha)$ reais. O acréscimo $c_0 \cdot \alpha$ (juro) é uma espécie de aluguel do dinheiro.

Sendo assim, raciocina o investidor, se eu resgatar meu capital depois de um semestre, terei direito a metade do juro (aluguel) anual, logo receberei $c_0\left(1 + \frac{\alpha}{2}\right)$ reais. Então reinvestirei esta soma por mais um semestre e, no final do ano, em vez de $c_0(1 + \alpha)$, vou receber $c_0\left(1 + \frac{\alpha}{2}\right)^2$, que é uma quantia maior. (Nosso investidor sabe que $\left(1 + \frac{\alpha}{2}\right)^2 > 1 + \alpha$, pela desigualdade de Bernoulli.) Pensando melhor, diz o investidor, posso resgatar e reinvestir meu capital mensalmente recebendo, no final de um ano, o total de $c_0\left(1 + \frac{\alpha}{12}\right)^{12}$.

Como o número $\alpha = k/100$ lhe é conhecido, o investidor, com auxílio da calculadora, verifica imediatamente que $\left(1 + \frac{\alpha}{2}\right)^2 < \left(1 + \frac{\alpha}{12}\right)^{12}$. Animado com o resultado, nosso ambicioso investidor imagina que, resgatando e reaplicando seu dinheiro num número n cada vez maior de intervalos de tempo iguais, poderá aumentar ilimitadamente seu capital.

Na verdade, fazendo o que imagina, no final do ano o investidor receberá o total acumulado igual a

$$c_0 \cdot \lim_{n\to\infty}\left(1 + \frac{\alpha}{n}\right)^n = c_0 \cdot e^{\alpha}.$$

Nosso personagem estava certo ao pensar que, para todo $n \in \mathbb{N}$ e todo $\alpha > 0$, se tem

$$\left(1 + \frac{\alpha}{n}\right)^n < \left(1 + \frac{\alpha}{n+1}\right)^{n+1}.$$

Mas, infelizmente, se enganou ao acreditar que a seqüência de termo geral $\left(1+\frac{\alpha}{n}\right)^n$ é ilimitada. Com efeito, todos esses termos são menores do que e^{α}.

Seja como for, ao conceber esse processo imaginário de resgatar e reinvestir a cada instante seu capital, nosso investidor foi conduzido à noção de juros compostos, acumulados continuamente.

O mesmo raciocínio é válido se considerarmos, para um número real arbitrário $t > 0$, o capital c_0 aplicado durante t anos, à mesma taxa α. Se tivéssemos juros simples, no final desses t anos o capital resultante seria $c_0(1+\alpha t)$. Dividindo o intervalo $[0, t]$ em n partes iguais, resgatando e reinvestindo n vezes, no final de t anos obteríamos $c_0\left(1+\frac{\alpha t}{n}\right)^n$ e, fazendo n crescer indefinidamente, chegamos a

$$c(t) = c_0 e^{\alpha t} = c_0 \cdot \lim_{n\to\infty}\left(1+\frac{\alpha t}{n}\right)^n$$

como o resultado da aplicação do capital c_0, durante t anos, a uma taxa de $\alpha = k/100$ ao ano, de juros compostos, acumulados continuamente.

Em particular, o capital de 1 real aplicado a uma taxa de 100% ao ano, com juros acumulados continuamente, gera no final de um ano um total de e reais.

Evidentemente, a expressão $f(t) = c \cdot e^{\alpha t}$ pode também ser escrita sob a forma $f(t) = c \cdot a^t$, onde $a = e^{\alpha}$, portanto $\alpha = \ell n\, a$. Ou, se houver preferência por uma determinada base b, pode-se sempre escrever $f(t) = c \cdot b^{\beta t}$, com $\beta = \frac{\alpha}{\ell n\, b}$. Às vezes é conveniente tomar a base 2, de modo que se tem $f(t) = c \cdot 2^{\beta t}$, onde $\beta = \alpha/\ell n\, 2$.

Matemáticos e cientistas que se utilizam da Matemática preferem geralmente escrever as funções do tipo exponencial sob a forma $f(x) = b \cdot e^{\alpha x}$, com a base e, porque esta expressão exibe explicitamente não apenas o valor inicial $b = f(0)$ como também o coeficiente α, que está intimamente ligado à taxa de crescimento de f, conforme mostraremos agora.

A *taxa de crescimento* de uma função f no intervalo de extremidades x, x + h é, por definição, o quociente

$$\frac{f(x+h)-f(x)}{h}.$$

Este quociente pode também ser interpretado como a inclinação da secante que liga os pontos $(x, f(x))$ e $(x+h, f(x+h))$ do gráfico de f.

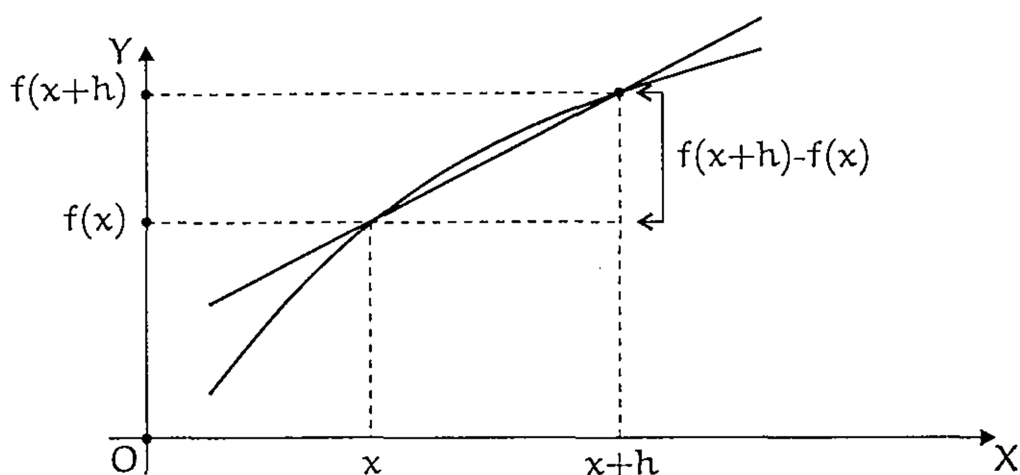


Figura 69

No caso particular da função $f(x) = be^{\alpha x}$, temos

$$\frac{f(x+h)-f(x)}{h} = be^{\alpha x}\frac{e^{\alpha h}-1}{h} = f(x)\cdot\frac{e^{\alpha h}-1}{n}.$$

Chama-se *derivada* da função f no ponto x ao limite da taxa $[f(x+h)-f(x)]/h$ quando h tende para zero. Este número, cujo significado é o de taxa instantânea de crescimento de f no ponto x, é representado por $f'(x)$. Ele é o número real cujos valores aproximados são os quocientes $[f(x+h)-f(x)]/h$ para valores muito pequenos de h. Geometricamente, a derivada $f'(x)$ é a inclinação da reta tangente ao gráfico da função f no ponto x.

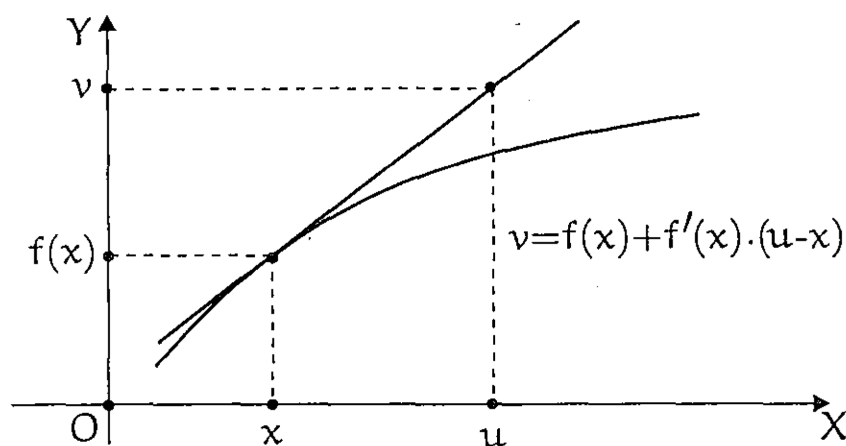


Figura 70

O sinal e o valor da derivada $f'(x)$ indicam a tendência da variação de f a partir do ponto x. Se $f'(x) > 0$ então $f(x+h) > f(x)$ para pequenos valores positivos de h. Se $f'(x) < 0$, tem-se, ao contrário, $f(x+h) < f(x)$ para h pequeno e positivo. Se $f'(x)$ é um número positivo grande, então f cresce rapidamente a partir de x. E assim por diante. A derivada é a noção fundamental do Cálculo Infinitesimal. Sua descoberta, há três séculos e meio, teve uma grande repercussão e provocou um progresso extraordinário na Ciência e em toda a civilização a partir daquela época.

Mostraremos agora que a derivada da função $f(x) = be^{\alpha x}$ é igual a $\alpha \cdot f(x)$. Noutras palavras, a taxa instantânea de crescimento de uma função do tipo exponencial é, em cada ponto x, proporcional ao valor da função naquele ponto. E o coeficiente α é precisamente o fator de proporcionalidade.

Assim, por exemplo, no caso do investimento, em que $c(t) = c_0 \cdot e^{\alpha t}$, se, a partir de um dado instante t_0, considerarmos um intervalo de tempo h muito pequeno, teremos aproximadamente $[c(t_0+h) - c(t_0)]/h \cong \alpha \cdot c(t_0)$, logo $c(t_0+h) - c(t_0) \cong c(t_0) \cdot \alpha h$.

Usando a interpretação geométrica do logaritmo natural, é fácil calcular a derivada da função $f(x) = b \cdot e^{\alpha x}$.

O ponto de partida consiste em mostrar que se tem

$$\lim_{h \to 0} \frac{e^h - 1}{h} = 1.$$

Para ver isto, lembramos que a faixa de hipérbole $H_1^{e^h}$ tem área igual a h. Esta faixa está compreendida entre um retângulo de área $(e^h - 1)/e^h$ e outro de área $e^h - 1$. Portanto

$$\frac{e^h - 1}{e^h} < h < e^h - 1.$$

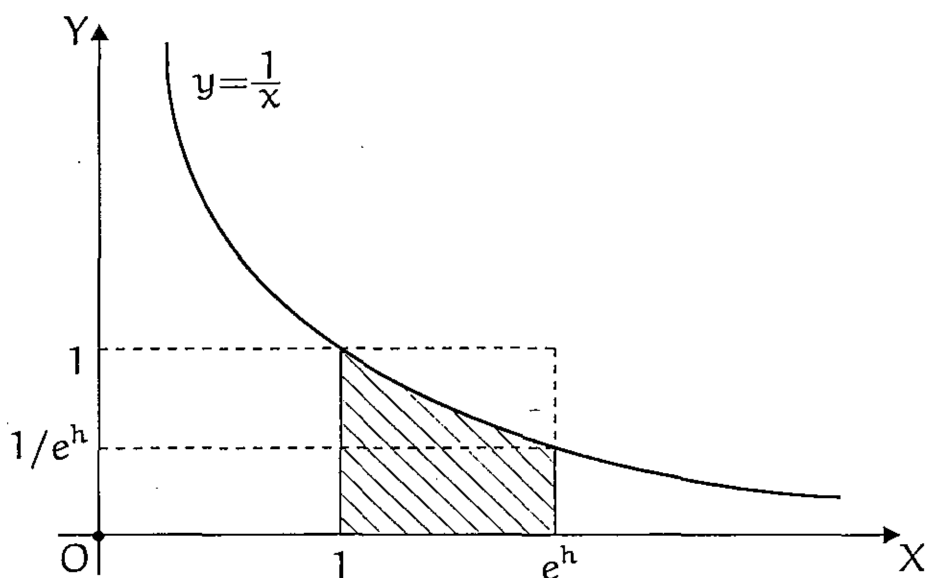


Figura 71

Aqui estamos supondo $h > 0$. Dividindo as duas desigualdades por $e^h - 1$, obtemos

$$\frac{1}{e^h} < \frac{h}{e^h - 1} < 1, \quad \text{para todo } h > 0.$$

Quando $h \to 0$, a potência e^h tende a 1. Segue-se das desigualdades acima que $\lim_{h\to 0} [h/(e^h - 1)] = 1$, logo

$$\lim_{h\to 0} \frac{e^h - 1}{h} = 1.$$

O caso em que $h \to 0$ por valores negativos se trata de modo análogo.

Agora é imediato ver que

$$\lim_{h\to 0} \frac{e^{x+h} - e^x}{h} = e^x \cdot \lim_{h\to 0} \frac{e^h - 1}{h} = e^x$$

e, mais geralmente,

$$\lim_{h\to 0} \frac{e^{\alpha(x+h)} - e^{x}}{h} = e^{\alpha x} \lim_{h\to 0} \frac{e^{\alpha h} - 1}{h} = \alpha \cdot e^{\alpha x} \cdot \lim_{h\to 0} \frac{e^{\alpha h} - 1}{\alpha h}.$$

Escrevendo $k = \alpha h$, vemos que $h \to 0 \Leftrightarrow k \to 0$. Portanto

$$\lim_{h\to 0} \frac{e^{\alpha(x+h)} - e^{\alpha x}}{h} = \alpha \cdot e^{\alpha x} \cdot \lim_{k\to 0} \frac{e^{k} - 1}{k} = \alpha \cdot e^{\alpha x}.$$

Isto conclui a demonstração de que a derivada da função $f(x) = e^{\alpha x}$ é $f'(x) = \alpha \cdot f(x)$, logo é proporcional ao valor $f(x)$ da função f, sendo α o fator de proporcionalidade.

É óbvio que o mesmo vale para uma função do tipo $f(x) = b \cdot e^{\alpha x}$.

11. Como Verificar que $f(x+h)/f(x)$ Depende Apenas de h

O teorema de caracterização das funções de tipo exponencial fornece um critério elegante e matematicamente simples para determinar quando uma bijeção crescente ou decrescente $f\colon \mathbb{R} \to \mathbb{R}^+$ é da forma $f(x) = b \cdot a^x$, ou seja, $f(x) = b \cdot e^{\alpha x}$.

Para aplicar esse critério em situações concretas é indispensável saber decidir, em cada caso específico, se $f(x+h)/f(x)$ independe de x ou não.

Fixando h como constante, isto equivale a indagar se $f(x + h)/f(x)$ é constante, isto é, se $f(x+h) = c \cdot f(x)$ para todo $x \in \mathbb{R}$, ou ainda, se $f(x+h)$ é uma função linear de $f(x)$. (Não de x!)

Escrevendo $y = f(x)$, $y' = f(x')$ e pondo $f(x+h) = \xi(y)$, $f(x'+h) = \xi(y')$, o Teorema Fundamental da Proporcionalidade diz que ξ é uma função linear de y se, e somente se, para quaisquer $y \in \mathbb{R}^+$ e $n \in \mathbb{N}$ tem-se a implicação:

$$y' = n \cdot y \quad \Rightarrow \quad \xi(y') = n \cdot \xi(y).$$

Em termos da função original f isto significa:

$$f(x') = n \cdot f(x) \quad \Rightarrow \quad f(x'+h) = n \cdot f(x+h).$$

A implicação acima é, portanto, o critério que nos permitirá decidir se a função f é ou não do tipo exponencial. Vejamos como ele funciona em alguns exemplos.

Capital a juros fixos. Aqui $c(t)$ é o capital no instante t, resultante da aplicação, a juros fixos acumulados continuamente, de um capital inicial $c_0 = c(0)$. Então $c(t+h)$ pode ser considerado como o capital resultante da aplicação da quantia inicial $c(t)$ durante o tempo h. Logo, aplicando o valor $n \cdot c(t) = c(t')$ obtém-se, após o período h, $n \cdot c(t+h)$. Portanto, $c(t') = n \cdot c(t) \Rightarrow c(t'+h) = n \cdot c(t+h)$. Segue-se que $c(t) = c_0 \cdot a^t$, onde $a = c(1)/c_0$, ou $c(t) = c_0 \cdot e^{\alpha t}$, onde $\alpha = \ell n\, a$. Como $c(t)$ é uma função crescente de t, tem-se $a > 1$, ou seja, $\alpha > 0$.

Desintegração radioativa. Neste exemplo, $m(t)$ é a massa, no instante t, de uma substância radioativa que no início da contagem do tempo era $m_0 = m(0)$. Assim, $m(t+h)$ é o que resta da massa $m(t)$ da substância radioativa depois de decorrido o intervalo de tempo h a partir do instante t. É claro que se observarmos a massa $m(t') = n \cdot m(t)$, $n \in \mathbb{N}$, após o mesmo tempo h, veremos que restou $m(t'+h) = n \cdot m(t+h)$. Portanto podemos assegurar que $m(t) = m_0 \cdot e^{\alpha t}$ para todo $t \in \mathbb{R}$. Neste caso, como $m(t)$ é função decrescente de t, temos $\alpha < 0$.

Concentração de uma solução. Este é o protótipo de uma situação que ocorre em diversas circunstâncias, inclusive a eliminação de substâncias na corrente sangüínea humana. Vamos considerar um caso bastante simples, que pode indicar como se tratam questões análogas.

Neste exemplo, temos um tanque de volume V, no qual se encontra uma salmoura (solução de sal em água), que se mantém homogênea mediante a ação permanente de um misturador. O tanque recebe um fluxo constante de água enquanto uma torneira escoa a salmoura em quantidade igual, a cada instante, ao volume de água que entrou no tanque.

Procura-se determinar a fórmula que exprime o volume $f(t)$ do

sal existente no tanque no momento t, portanto a taxa $f(t)/V$ que mede a concentração da solução naquele instante. Evidentemente, $f(t)$ é uma função decrescente de t.

Afirmamos que f é uma função de tipo exponencial. Com efeito, sejam $t, t' \in \mathbb{R}$ e $n \in \mathbb{N}$ tais que $f(t') = n \cdot f(t)$. (Observe que isto implica $t' \leqslant t$.) Fixemos arbitrariamente um intervalo de tempo h. Devemos mostrar que $f(t' + h) = n \cdot f(t + h)$. Para isto, imaginemos o tanque subdividido em n partes de igual volume V/n. Como a mistura é homogênea, em cada uma dessas partes a quantidade de sal nela contida no instante t' é igual a $f(t')/n$, ou seja igual a $f(t)$. O que ocorre em cada uma das subdivisões é, em escala $1/n$, o mesmo que ocorre no tanque inteiro. Logo, no instante $t' + h$, cada subdivisão vai conter o volume $f(t + h)$ de sal. No todo, vemos que o volume do sal contido no tanque inteiro no instante $t' + h$ é $f(t' + h) = n \cdot f(t + h)$.

Portanto, se b é o volume do sal contido no tanque no instante $t = 0$, a fórmula que exprime a quantidade de sal existente no tanque no tempo t é $f(t) = b \cdot a^t$, onde $a = f(1)/f(0) < 1$, ou seja, $f(t) = b \cdot e^{\alpha t}$, com $\alpha = \ell n\, a < 0$. Pela seção 10, temos $\alpha = f'(0)/b$, onde $f'(0) = \lim_{t \to 0}[f(t) - b]/t$. Seja v o volume de água que entra no tanque (igual ao da salmoura que sai) na unidade de tempo. Num tempo t, a água que entra é vt e, se t for muito pequeno, o sal que sai na salmoura é aproximadamente $(b/V)vt$. Logo $f(t) \approx b - (b/V)vt$ e

$$f'(0) = \lim_{t \to 0}[f(t) - b]/t = -bv/V.$$

Assim $\alpha = -v/V$ e, em qualquer instante t, a quantidade de sal no tanque é $f(t) = b.e^{-(v/V)t}$.

Exercícios

1. Com um lápis cuja ponta tem 0,02mm de espessura, deseja-se traçar o gráfico da função $f(x) = 2^x$. Até que distância à esquerda do eixo vertical pode-se ir sem que o gráfico atinja o eixo horizontal?

2. Dê exemplo de uma função crescente $f: \mathbb{R} \to \mathbb{R}^+$ tal que, para

todo $x \in \mathbb{R}$, a seqüência $f(x+1), f(x+2), \ldots, f(x+n), \ldots$ é uma progressão geométrica mas f não é do tipo $f(x) = b \cdot a^x$.

3. Dados $a > 0$ e $b > 0$, qual a propriedade da função exponencial que assegura a existência de $h \neq 0$ tal que $b^x = a^{x/h}$ para todo $x \in \mathbb{R}$? Mostre como obter o gráfico de $y = b^x$ a partir do gráfico de $y = a^x$. Use sua conclusão para traçar o gráfico de $y = \left(1/\sqrt[3]{4}\right)^x$ a partir do gráfico de $y = 2^x$.

4. Prove que uma função do tipo exponencial fica determinada quando se conhecem dois de seus valores. Mais precisamente, se $f(x) = b \cdot a^x$ e $F(x) = B \cdot A^x$ são tais que $f(x_1) = F(x_1)$ e $f(x_2) = F(x_2)$ com $x_1 \neq x_2$ então $a = A$ e $b = B$.

5. Dados $x_0 \neq 0$ e $y_0 > 0$ quaisquer, prove que existe $a > 0$ tal que $a^{x_0} = y_0$.

6. Dados $x_0 \neq x_1$ e y_0, y_1 não-nulos com o mesmo sinal, prove que existem $a > 0$ e b tais que $b \cdot a^{x_0} = y_0$ e $b \cdot a^{x_1} = y_1$.

7. A grandeza y se exprime como $y = b \cdot a^t$ em função do tempo t. Sejam d o acréscimo que se deve dar a t para que y dobre e m (meia-vida de y) o acréscimo de t necessário para que y se reduza à metade. Mostre que $m = -d$ e $y = b \cdot 2^{t/d}$, logo $d = \log_a 2 = 1/\log_2 a$.

8. Observações por longo tempo mostram que, após períodos de mesma duração, a população da terra fica multiplicda pelo mesmo fator. Sabendo que essa população era de 2,68 bilhões em 1956 e 3,78 bilhões em 1972, pede-se: (a) O tempo necessário para que a população da terra dobre de valor; (b) A população estimada para o ano 2012; (c) Em que ano a população da terra era de 1 bilhão.

9. Dê um argumento independente de observações para justificar que, em condições normais, a população da terra após o decurso de períodos iguais fica multiplicada pela mesma constante.

10. Resolva os exercícios do livro "Logaritmos", especialmente os do último capítulo.

Capítulo 9

Funções Trigonométricas

1. Introdução

As funções trigonométricas constituem um tema importante da Matemática, tanto por suas aplicações (que vão desde as mais elementares, no dia-a-dia, até as mais complexas, na Ciência e na alta Tecnologia) como pelo papel central que desempenham na Análise.

A Trigonometria teve seu início na antiguidade remota, quando se acreditava que os planetas descreviam órbitas circulares em redor da Terra, surgindo daí o interesse em relacionar o comprimento da corda de uma circunferência com o ângulo central por ela subtendido. Se c é o comprimento da corda, α é o ângulo e r o raio da circunferência então $c = 2r \operatorname{sen}(\alpha/2)$. Esta é a origem da palavra *seno*, que provém de uma tradução equivocada do árabe para o latim, quando se confundiu o termo *jiba* (corda) com *jaib* (dobra, cavidade, *sinus* em latim). [Cfr. "Meu Professor de Matemática", pág. 187.]

O objeto inicial da Trigonometria era o tradicional problema da resolução de triângulos, que consiste em determinar os seis elementos dessa figura (três lados e três ângulos) quando se conhecem três deles, sendo pelo menos um deles um lado.

Posteriormente, com a criação do Cálculo Infinitesimal, e do seu prolongamento que é a Análise Matemática, surgiu a necessidade de atribuir às noções de seno, cosseno e suas associadas tangente, cotangente, secante e cosecante, o status de função real de uma variável real. Assim, por exemplo, ao lado de $\cos \hat{A}$, o cos-

seno do ângulo $\widehat{A}$, tem-se também $\cos x$, o cosseno do número real x, isto é, a função $\cos\colon \mathbb{R} \to \mathbb{R}$. Analogamente, têm-se as funções sen, tg, cotg, sec e cossec, completando as *funções trigonométricas*.

Uma propriedade fundamental das funções trigonométricas é que elas são periódicas. Por isso são especialmente adaptadas para descrever os fenômenos de natureza periódica, oscilatória ou vibratória, os quais abundam no universo: movimento de planetas, som, corrente elétrica alternada, circulação do sangue, batimentos cardíacos, etc.

A importância das funções trigonométricas foi grandemente reforçada com a descoberta de Joseph Fourier, em 1822, de que toda função periódica (com ligeiras e naturais restrições) é uma soma (finita ou infinita) de funções do tipo $a \cos nx + b \operatorname{sen} nx$. Para que se tenha uma idéia da relevância deste fato, que deu origem à chamada Análise de Fourier, basta dizer que, segundo o banco de dados da revista "Mathematical Reviews", o nome mais citado nos títulos de trabalhos matemáticos nos últimos 50 anos é o de Fourier.

Como se sabe desde o ensino fundamental, num triângulo retângulo de hipotenusa a e ângulos agudos $\widehat{B}$, $\widehat{C}$, opostos respectivamente aos catetos b e c, têm-se as definições:

$$\cos \widehat{B} = \frac{c}{a} = \text{(cateto adjacente)} \div \text{(hipotenusa)},$$

$$\operatorname{sen} \widehat{B} = \frac{b}{a} = \text{(cateto oposto)} \div \text{(hipotenusa)},$$

$$\text{e, analogamente, } \cos \widehat{C} = \frac{b}{a}, \ \operatorname{sen} \widehat{C} = \frac{c}{a}.$$

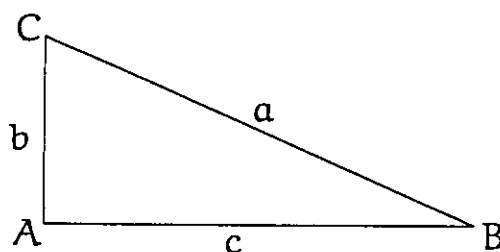


Figura 72

Estas relações definem o seno e o cosseno de um ângulo agudo qualquer pois todo ângulo agudo é um dos ângulos de um triângulo retângulo. É fundamental observar que $\cos \widehat{B}$ e $\operatorname{sen} \widehat{B}$ dependem apenas do ângulo $\widehat{B}$ mas não do tamanho do triângulo retângulo do qual $\widehat{B}$ é um dos ângulos agudos. Com efeito, dois quaisquer triângulos retângulos que tenham um ângulo agudo igual a $\widehat{B}$ são semelhantes.

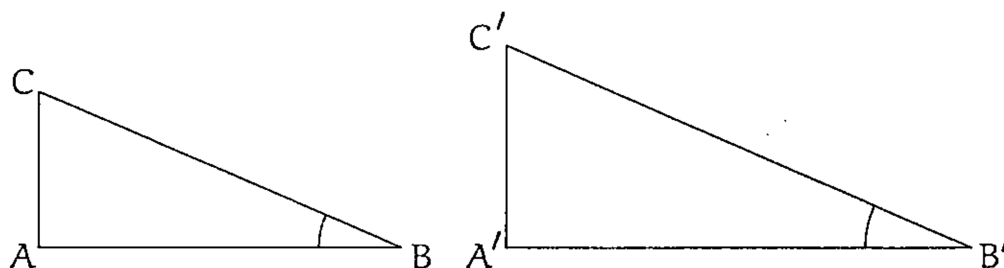


Figura 73

Se esses triângulos são ABC e $A'B'C'$, com $\widehat{B}' = \widehat{B}$. então a semelhança nos dá

$$\frac{b'}{a'} = \frac{b}{a}$$

e

$$\frac{c'}{a'} = \frac{c}{a},$$

logo

$$\operatorname{sen} \widehat{B}' = \operatorname{sen} \widehat{B} \quad \text{e} \quad \cos \widehat{B}' = \cos \widehat{B}.$$

Portanto o seno e o cosseno pertencem ao ângulo, e não ao eventual triângulo que o contém.

Assim, a semelhança de triângulos é a base de sustentação da Trigonometria. Se organizarmos uma tabela com os valores de $\cos \widehat{B}$ para todos os ângulos agudos $\widehat{B}$, a relação $c = a \cdot \cos \widehat{B}$ e o Teorema de Pitágoras

$$b = \sqrt{a^2 - c^2}$$

nos permitirão determinar os catetos b, c de um triângulo retângulo, uma vez conhecida a hipotenusa a e um dos ângulos agudos.

Mais geralmente, num triângulo ABC qualquer, a altura h, baixada do vértice C sobre o lado AB, tem a expressão $h = \overline{BC} \cdot \text{sen}\,\widehat{B}$. Esta simples fórmula exibe a eficiência da Trigonometria como instrumento de cálculo na Geometria, permitindo relacionar ângulos com comprimentos de segmentos.

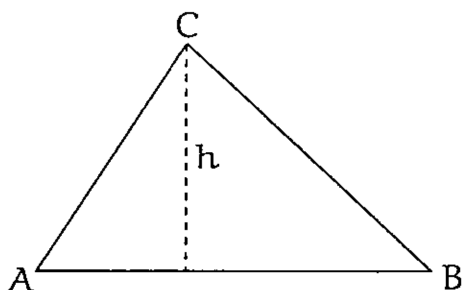


Figura 74

O Teorema de Pitágoras

$$a^2 = b^2 + c^2,$$

aplicado ao triângulo retângulo ABC, com $AB = c$, $AC = b$ e $BC = a$, nos mostra imediatamente que

$$(\cos \widehat{B})^2 + (\text{sen}\,\widehat{B})^2 = \frac{c^2}{a^2} + \frac{b^2}{a^2} = \frac{b^2 + c^2}{a^2} = \frac{a^2}{a^2} = 1.$$

É um costume tradicional, que convém adotar, escrever $\cos^2 \widehat{B}$ e $\text{sen}^2\,\widehat{B}$ em vez de $(\cos \widehat{B})^2$ e $(\text{sen}\,\widehat{B})^2$. A relação fundamental

$$\cos^2 \widehat{B} + \text{sen}^2\,\widehat{B} = 1$$

mostra que, a rigor, basta construir uma tabela de senos para ter a de cossenos, ou vice-versa.

É evidente, a partir da definição, que o cosseno de um ângulo agudo é igual ao seno do seu complemento e vice-versa. Daí a palavra "cosseno" (seno do complemento).

É claro que o seno e o cosseno de um ângulo agudo são números compreendidos entre 0 e 1.

Finalmente observamos que se A_1B_1 é a projeção ortogonal de um segmento de reta AB sobre um eixo então os comprimentos de

AB e A_1B_1 são relacionados pela fórmula $\overline{A_1B_1} = \overline{AB} \cdot \cos\alpha$, onde α é o ângulo de AB com o referido eixo.

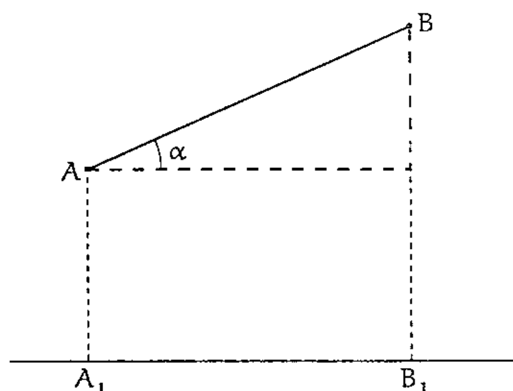


Figura 75

2. A Função de Euler e a Medida de Ângulos

A relação fundamental

$$\cos^2\alpha + \operatorname{sen}^2\alpha = 1$$

sugere que, para todo ângulo α, os números $\cos\alpha$ e $\operatorname{sen}\alpha$ são as coordenadas de um ponto da circunferência de raio 1 e centro na origem de $\mathbb{R}^2$.

Indicaremos com a notação C essa circunferência, que chamaremos de *circunferência unitária*, ou *círculo unitário*. Temos, portanto $C = \{(x,y) \in \mathbb{R}^2; x^2 + y^2 = 1\}$.

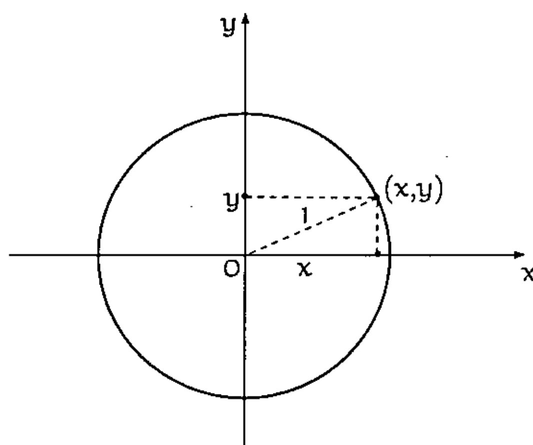


Figura 76

Observa-se que, para todo ponto $(x, y) \in C$, tem-se $-1 \leqslant x \leqslant 1$ e $-1 \leqslant y \leqslant 1$.

A fim de definir as funções $\cos: \mathbb{R} \to \mathbb{R}$ e $\text{sen}: \mathbb{R} \to \mathbb{R}$, devemos associar a cada número real t um ângulo e considerar o cosseno e o seno daquele ângulo. O número t desempenhará, portanto, o papel de medida do ângulo. Evidentemente, há diversas maneiras de se medir um ângulo, dependendo da unidade que se adota. Há duas unidades que se destacam: uma (o radiano) por ser, como veremos, a mais natural; outra (o grau) por ser tradicional há milênios, além de que muitos ângulos comumente encontrados têm por medida um número inteiro de graus.

A maneira natural de definir as funções trigonométricas tem como ponto de partida a função de Euler $E: \mathbb{R} \to C$, que faz corresponder a cada número real t o ponto $E(t) = (x, y)$ da circunferência unitária obtido do seguinte modo:

- $E(0) = (1, 0)$.

- se $t > 0$, percorremos sobre a circunferência C, a partir do ponto $(1, 0)$, um caminho de *comprimento* t, sempre andando no sentido positivo (contrário ao movimento dos ponteiros de um relógio comum, ou seja, o sentido que nos leva de $(1, 0)$ para $(0, 1)$ pelo caminho mais curto sobre C). O ponto final do caminho será chamado $E(t)$.

- se $t < 0$, $E(t)$ será a extremidade final de um caminho sobre C, de comprimento $|t|$, que parte do ponto $(1, 0)$ e percorre C sempre no sentido negativo (isto é, no sentido do movimento dos ponteiros de um relógio usual).

A função de Euler $E: \mathbb{R} \to C$ pode ser imaginada como o processo de enrolar a reta, identificada a um fio inextensível, sobre a circunferência C (pensada como um carretel) de modo que o ponto $0 \in \mathbb{R}$ caia sobre o ponto $(1, 0) \in C$.

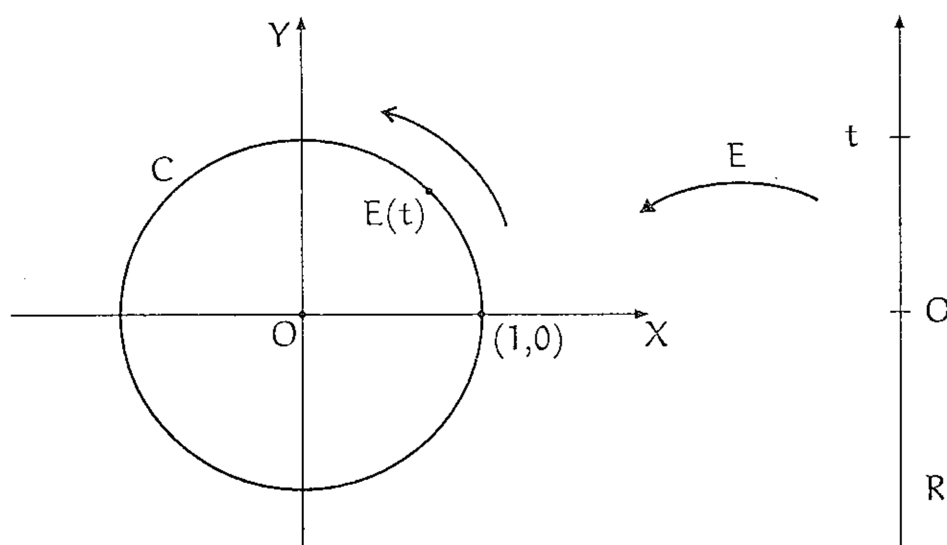


Figura 77

Cada vez que o ponto t descreve na reta um intervalo de comprimento ℓ, sua imagem $E(t)$ percorre sobre a circunferência C um arco de igual comprimento ℓ. Em particular, como a circunferência unitária C tem comprimento igual a 2π, quando o ponto t descreve um intervalo de comprimento 2π, sua imagem $E(t)$ dá uma volta completa sobre C, retornando ao ponto de partida. Assim sendo, para todo $t \in \mathbb{R}$, tem-se $E(t + 2\pi) = E(t)$ e, mais geralmente, para todo $k \in \mathbb{Z}$, tem-se $E(t + 2k\pi) = E(t)$, seja qual for $t \in \mathbb{R}$.

Reciprocamente, se $t < t'$ em $\mathbb{R}$ são tais que $E(t) = E(t')$ isto significa que, quando um ponto s da reta varia de t a t' sua imagem $E(s)$ se desloca sobre C, no sentido positivo, partindo de $E(t)$, dando um número inteiro k de voltas e retornando ao ponto de partida $E(t') = E(t)$. A distância total percorrida é igual a $2k\pi$, logo $t' = t + 2k\pi$, pois o comprimento do caminho percorrido por $E(s)$ é, por definição, igual à distância percorrida por s sobre a reta $\mathbb{R}$.

Resumindo: tem-se $E(t') = E(t)$ se, e somente se, $t' = t + 2k\pi$, com $k \in \mathbb{Z}$. (Quando $t' > t$, vale $k \in \mathbb{N}$; quando $t' < t$ tem-se $k < 0$.)

Escrevamos $A = (1, 0)$ e $O = (0, 0)$. Para cada $t \in \mathbb{R}$, ponhamos $B = E(t)$. Diz-se neste caso que o ângulo $A\hat{O}B$ *mede* t *radianos*.

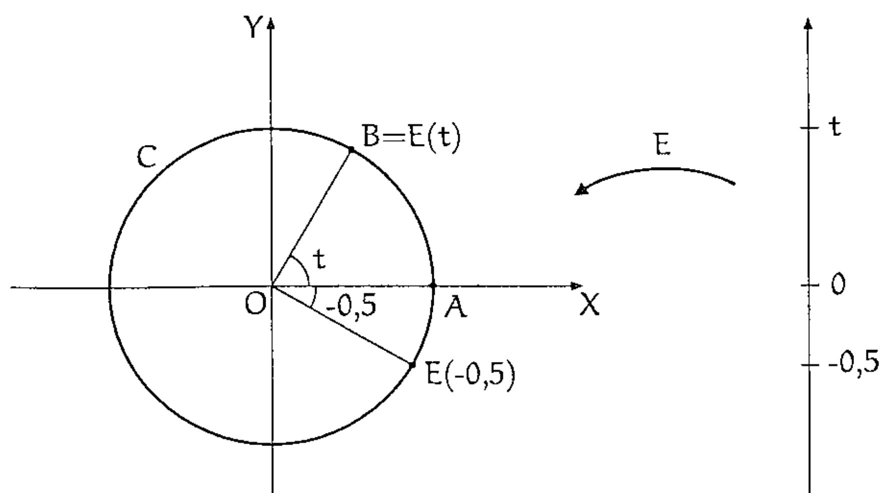


Figura 78

Esta definição sugere uma série de observações.

- Pode-se ter $B = E(t)$ com $t < 0$. Portanto esta forma de medida é *orientada*: é permitido a um ângulo ter medida negativa.
- A medida do ângulo $A\hat{O}B$ é determinada apenas a menos de um múltiplo inteiro de 2π, pois $B = E(t)$ implica $B = E(t+2k\pi)$ para todo $k \in \mathbb{Z}$. Assim, por exemplo, o ângulo de 1 radiano é também um ângulo de $1 - 2\pi$ radianos. De um modo mais geral, se $B = E(t)$ então $B = E(t - 2\pi)$ pois há dois arcos que vão de $A = (1, 0)$ até B; um de comprimento $|t|$ e outro de comprimento $|t - 2\pi|$.

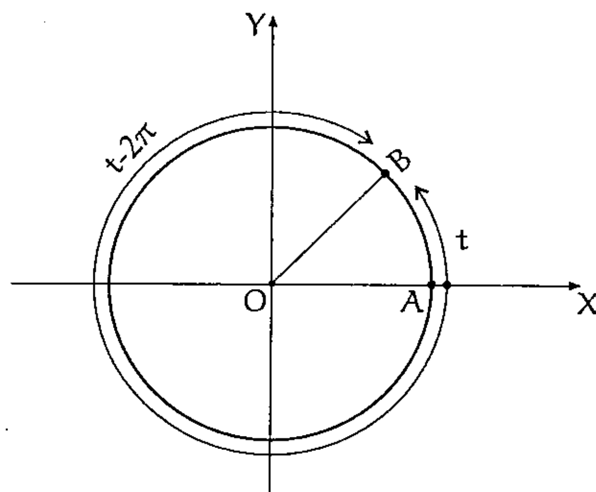


Figura 79

- De acordo com esta definição, o ângulo $A\hat{O}B$ mede 1 radiano se, e somente se, o arco $\widehat{AB}$ da circunferência C, por ele subtendido, tem comprimento igual a 1, isto é, igual ao raio da circunferência. Mais geralmente, numa circunferência de raio r, a medida de um ângulo central em radianos é igual a ℓ/r, onde ℓ é o comprimento do arco subtendido por esse ângulo.
- A medida do ângulo $A\hat{O}B$ em radianos também pode ser expressa como $2a/r^2$, em termos da área a do setor circular AOB e do raio r.

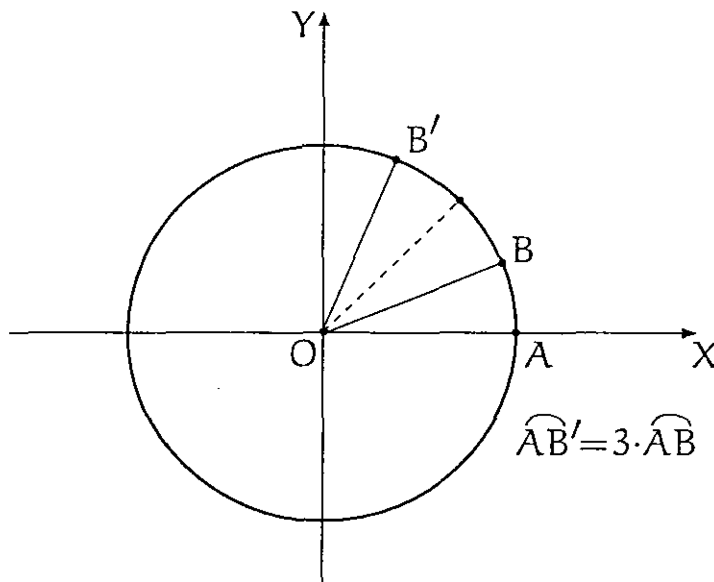


Figura 80

Com efeito, a área a do setor circular AOB é uma função crescente do comprimento ℓ do arco $\widehat{AB}$. Como se vê facilmente, se o arco $\widehat{AB'}$ tem comprimento n vezes maior do que o arco $\widehat{AB}$ (onde $n \in \mathbb{N}$) então a área do setor AOB' é igual a n vezes a área de AOB. Segue-se então do Teorema Fundamental da Proporcionalidade que a área a é uma função linear do comprimento ℓ: $a = c \cdot \ell$, onde c é uma constante. Para determinar o valor de c, basta observar que, quando o setor é todo o círculo (de raio r), o arco correspondente é toda a circunferência. Tem-se então $a = \pi r^2$ e $\ell = 2\pi r$. Logo $\pi r^2 = c \cdot 2\pi r$, donde $c = \frac{r}{2}$.

Portanto a área a do setor AOB se relaciona com o comprimento ℓ do arco $\overset{\frown}{AB}$ pela igualdade $a = \ell r/2$.

Segue-se que

$$\frac{\ell}{r} = \frac{2a}{r^2}.$$

Como ℓ/r é a medida do ângulo $A\hat{O}B$ em radianos, concluímos daí que esta medida também vale $2a/r^2$, onde a é a área do setor AOB e r é o raio do círculo.

Podíamos também ter definido uma função $G\colon \mathbb{R} \to C$ pondo ainda $G(0) = (1,0)$ e estipulando que, para $s > 0$, $G(s)$ fosse o ponto da circunferência unitária obtido a partir do ponto $(1,0)$ quando se percorre, ao longo de C, no sentido positivo, um caminho de comprimento $\frac{2\pi}{360}s$. E, para $s < 0$, $G(s)$ seria definido de forma análoga, com o percurso no sentido negativo de C.

A função $G\colon \mathbb{R} \to C$ tem propriedades semelhantes às de E, pois

$$G(t) = E(\frac{2\pi}{360}t)$$

para todo $t \in \mathbb{R}$. Em particular, $G(t') = G(t)$ se, e somente se, $t' = t + 360k$, onde $k \in \mathbb{Z}$.

Se $A = (1,0)$, $0 = (0,0)$ e $B = G(s)$, diz-se que o ângulo $A\hat{O}B$ mede s *graus*. O ângulo $A\hat{O}B$ mede 1 grau quando $B = G(1)$, ou seja, quando o arco $\overset{\frown}{AB}$ tem comprimento igual a $2\pi/360$. Noutras palavras, o ângulo de 1 grau é aquele que subtende um arco igual a $1/360$ da circunferência.

Escreve-se 1 grau $= 1°$ e 1 radiano $= 1$ rad.

Como a circunferência inteira tem 2π radianos e 360 graus, segue-se que 2π rad $= 360°$, ou seja,

$$1 \text{ rad} = (\frac{360}{2\pi})° = 57,3 \text{ graus}.$$

É bom ter em mente relações como $180° = \pi$ rad, $90° = \frac{\pi}{2}$ rad, etc.

As figuras abaixo deixam claro que se $E(t) = (x,y)$ então $E(t+\pi) = (-x,-y)$, $E(t+\frac{\pi}{2}) = (-y,x)$, $E(-t) = (x,-y)$, $E(\frac{\pi}{2}-t) = (y,x)$

e $E(\pi - t) = (-x, y)$.

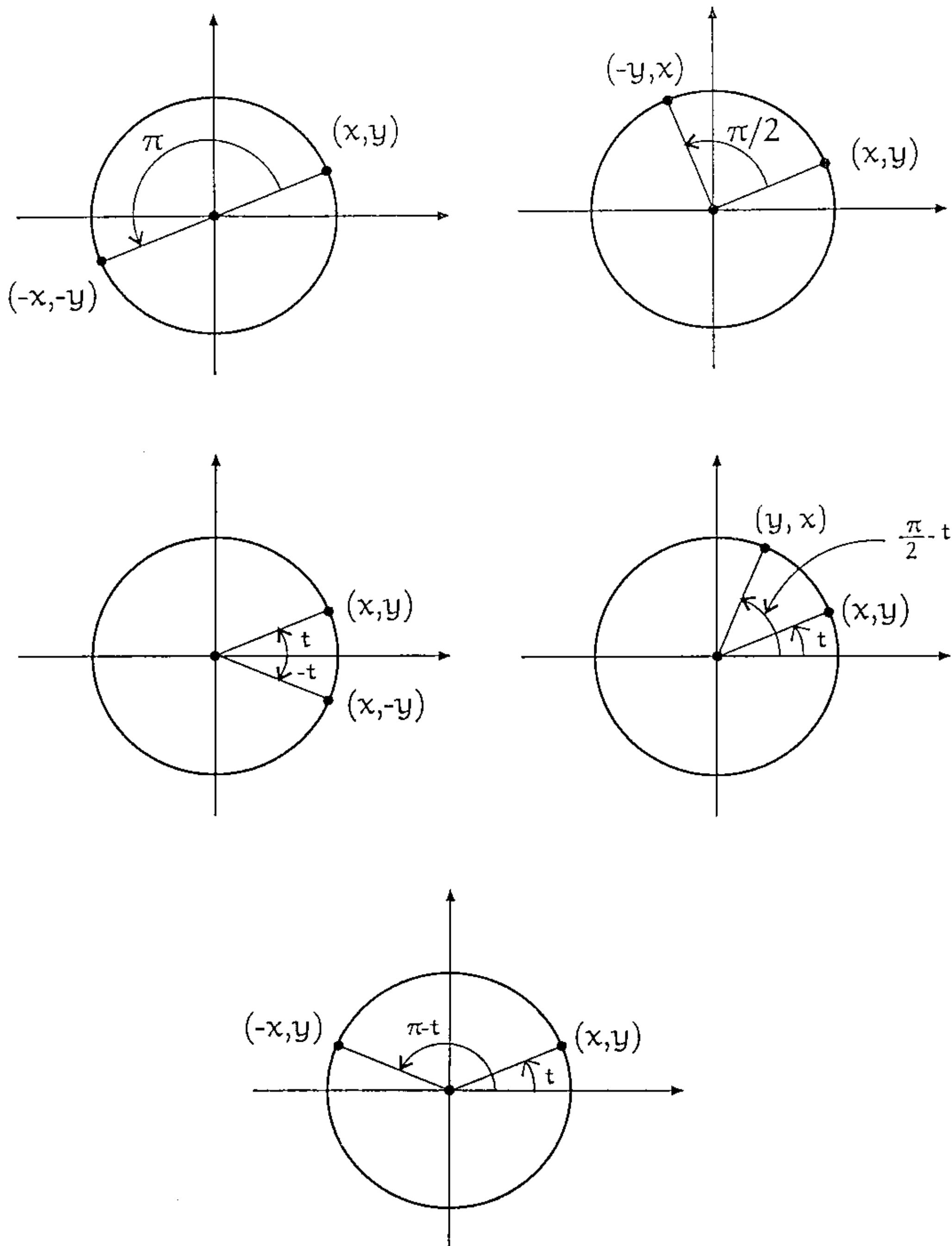


Figura 81

Estas relações exprimem certas simetrias da função de Euler $E: \mathbb{R} \to C$, que se traduzem em propriedades das funções seno e cosseno, como veremos a seguir.

3. As Funções Trigonométricas

As funções $\cos: \mathbb{R} \to \mathbb{R}$ e $\text{sen}: \mathbb{R} \to \mathbb{R}$, chamadas *função cosseno* e *função seno* respectivamente, são definidas pondo-se, para cada $t \in \mathbb{R}$:

$$E(t) = (\cos t, \text{sen } t).$$

Noutras palavras, $x = \cos t$ e $y = \text{sen } t$ são respectivamente a abcissa e a ordenada do ponto $E(t)$ da circunferência unitária.

Segue-se imediatamente desta definição que vale, para todo $t \in \mathbb{R}$, a relação fundamental

$$\cos^2 t + \text{sen}^2 t = 1.$$

Uma função $f: \mathbb{R} \to \mathbb{R}$ chama-se *periódica* quando existe um número $T \neq 0$ tal que $f(t+T) = f(t)$ para todo $t \in \mathbb{R}$. Se isto ocorre, então $f(t+kT) = f(t)$ para todo $t \in \mathbb{R}$ e todo $k \in \mathbb{Z}$. O menor número $T > 0$ tal que $f(t+T) = f(t)$ para todo $t \in \mathbb{R}$ chama-se o *período* da função f. As funções seno e cosseno são periódicas, de período 2π.

Diz-se ainda que a função $f: \mathbb{R} \to \mathbb{R}$ é *par* quando se tem $f(-t) = f(t)$ para todo $t \in \mathbb{R}$. Se se tem $f(-t) = -f(t)$ para todo $t \in \mathbb{R}$, a função f chama-se *ímpar*.

Exemplo. Seja $f: \mathbb{R} \to \mathbb{R}$ a função *dente-de-serra*, assim definida: $f(k) = 0$ se $k \in \mathbb{Z}$ e $f(k+\alpha) = \alpha$ quando $0 \leqslant \alpha < 1$ e $k \in \mathbb{Z}$. A função f é periódica, com período 1, mas não é par nem ímpar. Por outro lado, a função $g: \mathbb{R} \to \mathbb{R}$, onde $g(t) = t^n$ (com $n \in \mathbb{N}$) é par se n é um número par e é uma função ímpar quando n é um número ímpar.

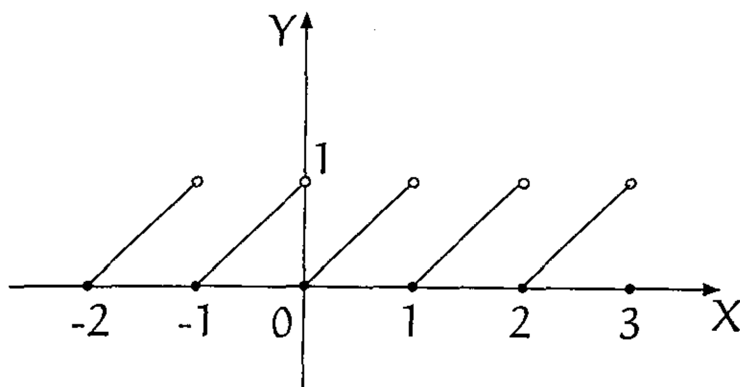


Figura 82

Para todo $t \in \mathbb{R}$, temos

$$E(t) = (\cos t, \operatorname{sen} t)$$

e

$$E(-t) = (\cos(-t), \operatorname{sen}(-t)).$$

Mas, como vimos no fim da seção anterior, quando $E(t) = (x, y)$ tem-se $E(-t) = (x, -y)$. Isto significa que

$$\cos(-t) = \cos t \text{ e } \operatorname{sen}(-t) = -\operatorname{sen} t$$

para todo $t \in \mathbb{R}$.

Assim, cosseno é uma função par e seno é uma função ímpar. De modo análogo, as outras quatro relações estabelecidas no final da seção anterior mostram que, para todo $t \in \mathbb{R}$, valem:

$$\begin{aligned}
&\cos(t+\pi) = -\cos t, \quad \operatorname{sen}(t+\pi) = -\operatorname{sen} t,\\
&\cos\left(t+\frac{\pi}{2}\right) = -\operatorname{sen} t, \quad \operatorname{sen}\left(t+\frac{\pi}{2}\right) = \cos t,\\
&\cos\left(\frac{\pi}{2}-t\right) = \operatorname{sen} t, \quad \operatorname{sen}\left(\frac{\pi}{2}-t\right) = \cos t,\\
&\cos(\pi - t) = -\cos t, \quad \operatorname{sen}(\pi - t) = \operatorname{sen} t.
\end{aligned}$$

As figuras mostram os gráficos de $y = \cos x$ e $y = \operatorname{sen} x$.

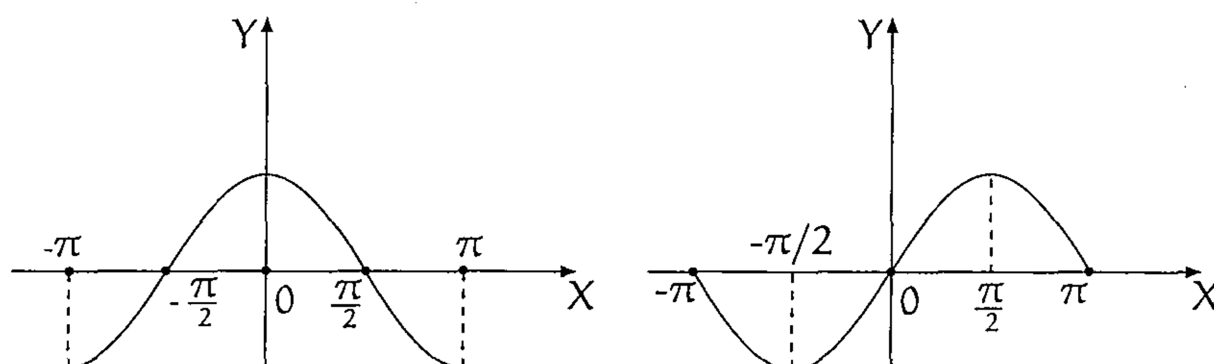


Figura 83

Alguns valores particulares das funções seno e cosseno podem ser obtidos mediante argumentos geométricos, alguns dos quais são interessantes exercícios, especialmente quando se usam as fórmulas de adição, que estabeleceremos a seguir. Do ponto de

vista numérico, entretanto, é claro que o modo mais eficiente de obter os valores dessas funções é usar uma calculadora, principalmente uma que opere com radianos e com graus.

Independentemente de calculadoras, é muito conveniente que se saiba, sem pensar muito, quais os valores de t que satisfazem as equações

$$\operatorname{sen} t = 0, \cos t = 0,$$
$$\operatorname{sen} t = 1, \cos t = 1,$$
$$\operatorname{sen} t = -1, \cos t = -1,$$
$$\operatorname{sen} t = \cos t,$$
$$\operatorname{sen} t = \frac{1}{2}, \cos t = \frac{1}{2}$$

e outras semelhantes.

Para interessantes exemplos, exercícios e um tratamento bastante completo dos assuntos aqui abordados, veja-se o livro "Trigonometria e Números Complexos", da Coleção do Professor de Matemática (SBM).

Das funções seno e cosseno derivam as outras funções trigonométricas, a saber $\operatorname{tg} x = \operatorname{sen} x/\cos x$, $\operatorname{cotg} x = \cos x/\operatorname{sen} x$, $\sec x = 1/\cos x$ e $\operatorname{cossec} x = 1/\operatorname{sen} x$. Destas funções (chamadas tangente, cotangente, secante e cossecante), a mais importante é a primeira. Cumpre observar que tais funções, sendo definidas por meio de quocientes, têm seus domínios restritos aos números reais para os quais o denominador é diferente de zero.

Assim, por exemplo, a função tangente, dada pela expressão $\operatorname{tg} x = \operatorname{sen} x/\cos x$, tem como domínio o conjunto dos números reais que não são múltiplos ímpares de $\pi/2$ pois $\cos x = 0$ se, e somente se, $x = (2k+1)\frac{\pi}{2} = k + \frac{\pi}{2}$ onde $k \in \mathbb{Z}$.

Assim, o domínio da função $x \mapsto \operatorname{tg} x$ é formado pela reunião dos intervalos abertos $\left(k\pi - \frac{\pi}{2}, k\pi + \frac{\pi}{2}\right)$, para todo $k \in \mathbb{Z}$.

Em cada um desses intervalos [por exemplo $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$] a função tangente é crescente e, na realidade, $x \mapsto \operatorname{tg} x$ é uma correspondência biunívoca entre um intervalo aberto de comprimento π e a reta

inteira $\mathbb{R}$.

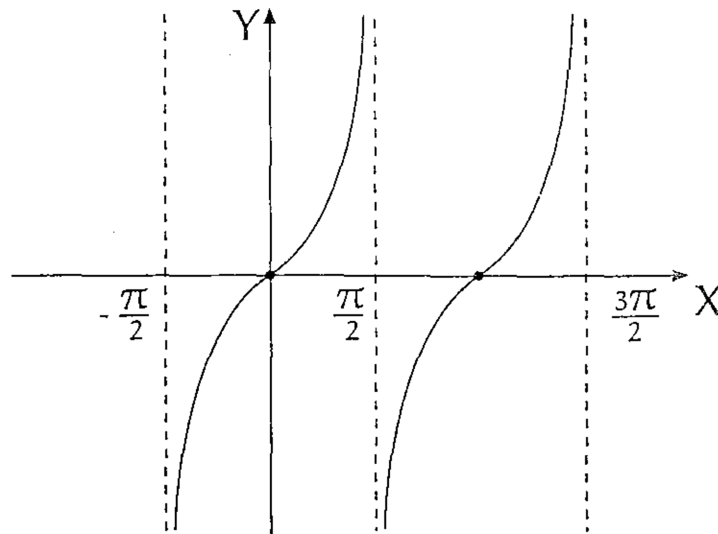


Figura 84

A função tangente, embora não esteja definida para todo número real $\mathbb{R}$, pode ser considerada como uma função periódica, de período π, pois π é o menor número real positivo tal que $\text{tg}(x+\pi) = \text{tg}\,x$ para todo x no domínio da função.

A restrição da função tangente ao intervalo $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, sendo uma correspondência biunívoca $\text{tg}\colon \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$, possui uma função inversa, chamada *arco tangente*, indicada com a notação $\text{arctg}\colon \mathbb{R} \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, a qual é uma correspondência biunívoca de domínio $\mathbb{R}$ e imagem igual ao intervalo aberto $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

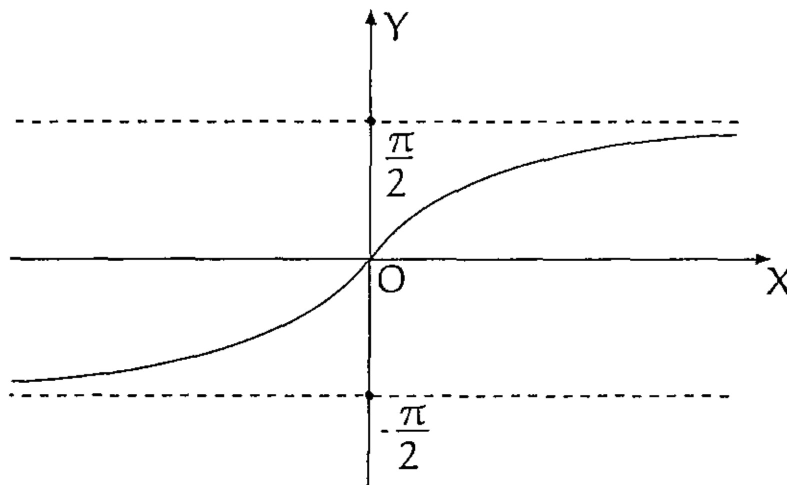


Figura 85

Para todo ponto $P = (x, y)$ em $\mathbb{R}^2$, com $x \neq 0$, se α é o ângulo do semi-eixo positivo $\overrightarrow{OX}$ com a semi-reta OP então

$$\frac{y}{x} = \operatorname{tg} \alpha.$$

Isto é verdadeiro, por definição, quando P está sobre a circunferência unitária e vale também no caso geral por semelhança de triângulos.

Segue-se daí que se $y = ax + b$ é uma reta não-vertical, o coeficiente a é a tangente do ângulo α que o semi-eixo positivo $\overrightarrow{OX}$ faz com essa reta. Com efeito, se tomarmos $x_1 \neq x_2$ e pusermos

$$y_1 = ax_1 + b,$$

$$y_2 = ax_2 + b,$$

teremos

$$a = \frac{y_2 - y_1}{x_2 - x_1} = \operatorname{tg} \alpha.$$

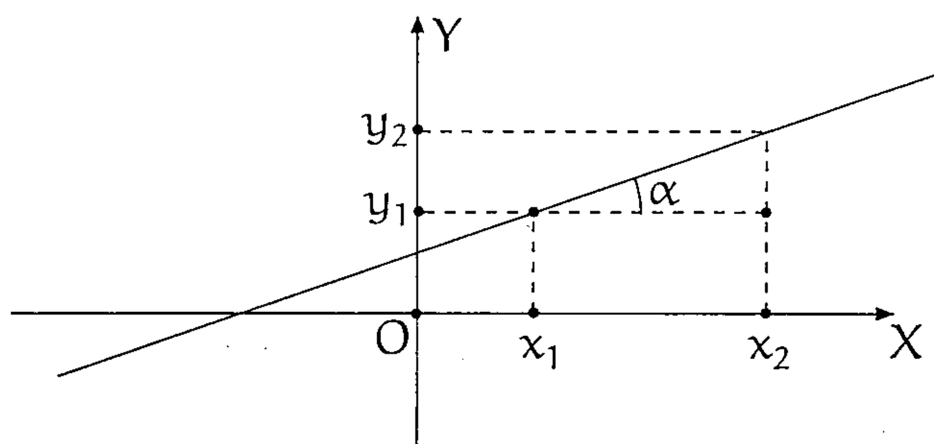


Figura 86

4. As Fórmulas de Adição

As fórmulas clássicas que exprimem $\cos(\alpha + \beta)$ e $\operatorname{sen}(\alpha + \beta)$ em termos de $\cos \alpha$, $\cos \beta$, $\operatorname{sen} \alpha$ e $\operatorname{sen} \beta$ podem ser demonstradas de vários modos. (Vide "Trigonometria e Números Complexos", já citado.) Daremos aqui a prova que nos parece a mais direta.

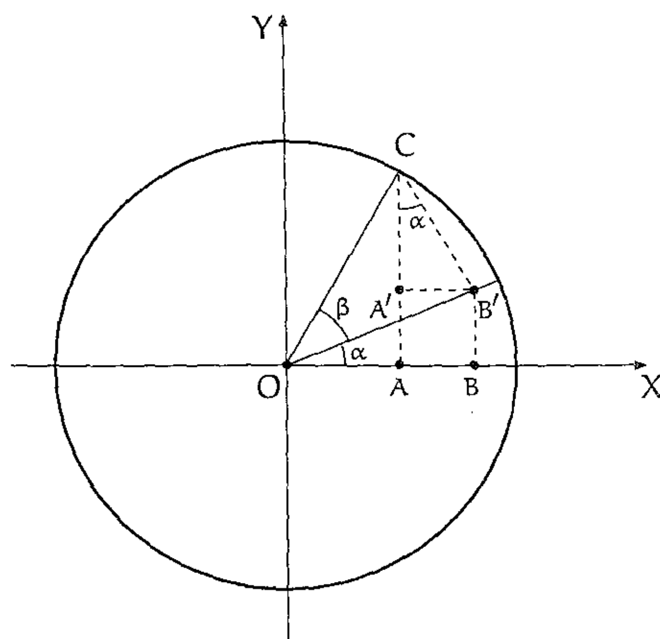


Figura 87

Na figura, onde $CB' \perp OB'$, temos

$$\overline{OA} = \cos(\alpha + \beta),$$

$$\overline{OB'} = \cos\beta,$$

$$\overline{B'C} = \operatorname{sen}\beta,$$

$$\overline{AB} = \overline{A'B'} = \operatorname{sen}\alpha \cdot \operatorname{sen}\beta \quad \text{e}$$

$$\overline{OB} = \cos\alpha \cdot \cos\beta.$$

Logo

$$\overline{OA} = \overline{OB} - \overline{AB} = \cos\alpha \cdot \cos\beta - \operatorname{sen}\alpha \cdot \operatorname{sen}\beta.$$

Noutras palavras,

$$\cos(\alpha + \beta) = \cos\alpha \cdot \cos\beta - \operatorname{sen}\alpha \cdot \operatorname{sen}\beta.$$

Tomando $-\beta$ em vez de β na fórmula acima, como $\cos(-\beta) = \cos\beta$ e $\operatorname{sen}(-\beta) = -\operatorname{sen}\beta$, obtemos:

$$\cos(\alpha - \beta) = \cos\alpha \cdot \cos\beta + \operatorname{sen}\alpha \cdot \operatorname{sen}\beta.$$

Além disso, como

$$\operatorname{sen}\left(\frac{\pi}{2} + t\right) = \cos t$$

e

$$\cos\left(\frac{\pi}{2} + t\right) = -\operatorname{sen} t,$$

a fórmula de $\cos(\alpha + \beta)$ nos dá também:

$$\begin{aligned}\operatorname{sen}(\alpha+\beta) &= -\cos\left(\frac{\pi}{2} + \alpha + \beta\right)\\ &= -\cos\left(\frac{\pi}{2} + \alpha\right)\cos\beta + \operatorname{sen}\left(\frac{\pi}{2} + \alpha\right)\operatorname{sen}\beta,\end{aligned}$$

ou seja,

$$\operatorname{sen}(\alpha+\beta) = \operatorname{sen}\alpha \cdot \cos\beta + \operatorname{sen}\beta \cdot \cos\alpha.$$

Daí resulta imediatamente que

$$\operatorname{sen}(\alpha-\beta) = \operatorname{sen}\alpha \cdot \cos\beta - \operatorname{sen}\beta \cdot \cos\alpha.$$

As fórmulas para o seno e o cosseno do arco duplo são conseqüências diretas:

$$\cos 2\alpha = \cos^2\alpha - \operatorname{sen}^2\alpha \quad \text{e} \quad \operatorname{sen} 2\alpha = 2\operatorname{sen}\alpha \cdot \cos\alpha.$$

Como aplicação das fórmulas de adição, mostraremos como determinar as coordenadas do ponto $A' = (x', y')$, obtido do ponto $A = (x, y)$ por meio da rotação de ângulo θ em torno da origem de $\mathbb{R}^2$.

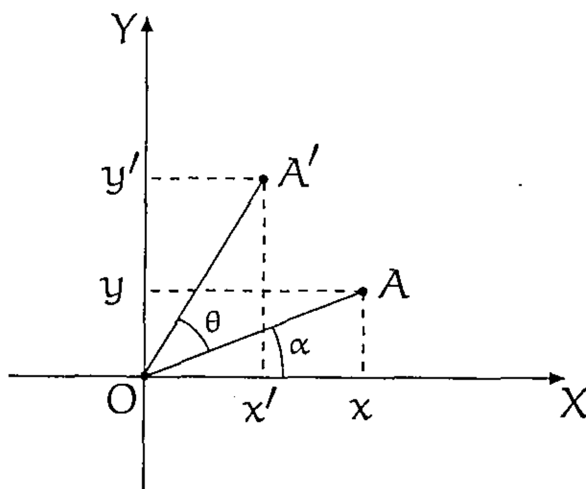


Figura 88

Chamemos de α o ângulo do eixo OX com o segmento OA e escrevamos $r = \overline{OA}$. Então $r = \overline{OA'}$ e se tem

$$x = r \cdot \cos\alpha, \quad y = r \cdot \operatorname{sen}\alpha, \quad x' = r \cdot \cos(\alpha + \theta), \quad y' = r \cdot \operatorname{sen}(\alpha + \theta).$$

As fórmulas de adição fornecem

$$x' = r\cos\alpha \cdot \cos\theta - r\operatorname{sen}\alpha \cdot \operatorname{sen}\theta = x\cos\theta - y\operatorname{sen}\theta,$$
$$y' = r\cos\alpha \cdot \operatorname{sen}\theta + r\operatorname{sen}\alpha \cdot \cos\theta = x\operatorname{sen}\theta + y\cos\theta.$$

Portanto a rotação de ângulo θ em torno da origem é a função (transformação) $T: \mathbb{R}^2 \to \mathbb{R}^2$ definida por

$$T(x, y) = (x\cos\theta - y\operatorname{sen}\theta, x\operatorname{sen}\theta + y\cos\theta).$$

Outra aplicação interessante das fórmulas de adição consiste em mostrar que $\cos\alpha$ e $\operatorname{sen}\alpha$ se exprimem como funções racionais de $\operatorname{tg}\frac{\alpha}{2}$, fato que está intimamente ligado com a parametrização racional da circunferência unitária C, conforme veremos agora.

É um fato bastante conhecido, e muito fácil de constatar, que para todo número real x vale a igualdade

$$\left(\frac{1 - x^2}{1 + x^2}\right)^2 + \left(\frac{2x}{1 + x^2}\right)^2 = 1.$$

Isto significa que, para todo $x \in \mathbb{R}$, os números dentro dos parênteses acima são respectivamente a abcissa e a ordenada de um ponto da circunferência unitária C, isto é, são o cosseno e o seno de um ângulo β. Além disso, todo número real x é a tangente de um (único) ângulo $\alpha \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Logo a igualdade acima significa que, para cada um desses valores de α, existe um β tal que

$$\frac{1 - \operatorname{tg}^2\alpha}{1 + \operatorname{tg}^2\alpha} = \cos\beta \quad \text{e} \quad \frac{2\operatorname{tg}\alpha}{1 + \operatorname{tg}^2\alpha} = \operatorname{sen}\beta.$$

É fácil mostrar que $\beta = 2\alpha$ usando as fórmulas de $\cos 2\alpha$ e $\operatorname{sen} 2\alpha$. Basta substituir $\operatorname{tg}\alpha$ por $\operatorname{sen}\alpha / \cos\alpha$ no primeiro membro destas

igualdades e fazer as simplificações óbvias para ver que

$$\frac{1-\operatorname{tg}^2\alpha}{1+\operatorname{tg}^2\alpha}=\cos 2\alpha \quad \text{e} \quad \frac{2\operatorname{tg}\alpha}{1+\operatorname{tg}^2\alpha}=\operatorname{sen} 2\alpha.$$

Equivalentemente:

$$\cos\alpha=\frac{1-\operatorname{tg}^2\frac{\alpha}{2}}{1+\operatorname{tg}^2\frac{\alpha}{2}}, \quad \operatorname{sen}\alpha=\frac{2\operatorname{tg}\frac{\alpha}{2}}{1+\operatorname{tg}^2\frac{\alpha}{2}}.$$

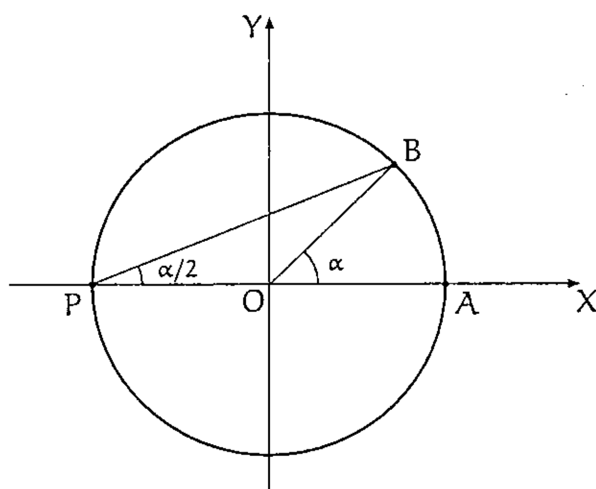


Figura 89

Dado o ponto arbitrário $B = (\cos\alpha, \operatorname{sen}\alpha)$ da circunferência unitária, como o ângulo inscrito $A\widehat{P}B$ é a metade do ângulo central $\alpha = A\widehat{O}B$ que subtende o mesmo arco $\overset{\frown}{AB}$, vemos que $\operatorname{tg}\frac{\alpha}{2}$ é a inclinação da reta PB, onde $P = (-1, 0)$. Mantendo o ponto P fixo e fazendo $\frac{\alpha}{2}$ variar em $(-\pi/2, +\pi/2)$, cada semi-reta de inclinação igual a $\operatorname{tg}\frac{\alpha}{2}$ corta a circunferência unitária num único ponto $B = (\cos\alpha, \operatorname{sen}\alpha)$. Todos os pontos da circunferência podem ser obtidos assim, menos o próprio ponto P.

A correspondência

$$x \mapsto \left(\frac{1-x^2}{1+x^2}, \frac{2x}{1+x^2}\right)$$

é uma parametrização racional de C. Para todo $x \in \mathbb{Q}$, o ponto que lhe corresponde tem ambas as coordenadas racionais.

5. A Lei dos Cossenos e a Lei dos Senos

Dado o triângulo ABC, sejam a, b, c as medidas dos lados BC, AC e AB respectivamente. Seja ainda $h = \overline{AP}$ a altura baixada de A sobre o lado BC. Há duas possibilidades, ilustradas nas figuras, conforme o ponto P pertença ao segmento BC ou esteja sobre seu prolongamento.

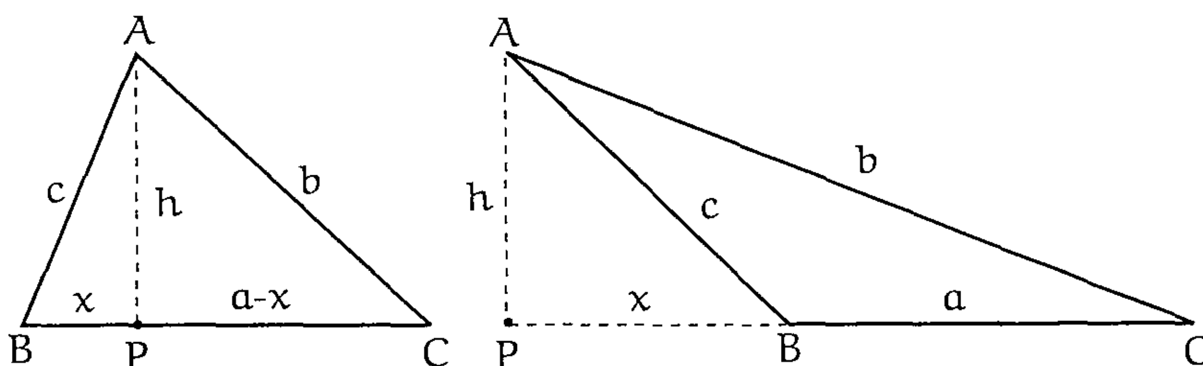


Figura 90

No primeiro caso, seja $x = \overline{BP} = c \cdot \cos\widehat{B}$. O Teorema de Pitágoras aplicado aos triângulos ABP e APC fornece as igualdades

$$\begin{aligned} c^2 &= h^2 + x^2, \\ b^2 &= h^2 + (a-x)^2 = h^2 + x^2 + a^2 - 2ax \\ &= h^2 + x^2 + a^2 - 2ac \cdot \cos\widehat{B}. \end{aligned}$$

Comparando estas igualdades obtemos

$$b^2 = a^2 + c^2 - 2ac \cdot \cos\widehat{B}.$$

No segundo caso, $x = \overline{BP} = c \cdot \cos(\pi - \widehat{B}) = -c \cdot \cos\widehat{B}$. (Note que $\cos\widehat{B} < 0$, logo $-c \cdot \cos\widehat{B}$ é positivo.) Novamente Pitágoras, aplicado aos triângulos APB e APC nos dá:

$$\begin{aligned} c^2 &= h^2 + x^2, \\ b^2 &= h^2 + (a+x)^2 = h^2 + x^2 + a^2 + 2ax \\ &= h^2 + x^2 + a^2 - 2ac \cdot \cos\widehat{B}. \end{aligned}$$

Daí resulta, como antes, que

$$b^2 = a^2 + c^2 - 2ac \cdot \cos \widehat{B}.$$

Portanto a igualdade vale em qualquer caso. Ela é a *lei dos cossenos*, da qual o Teorema de Pitágoras é um caso particular, que se tem quando $\widehat{B}$ é um ângulo reto.

Evidentemente, tem-se também

$$a^2 = b^2 + c^2 - 2bc \cdot \cos \widehat{A}$$

e

$$c^2 = a^2 + b^2 - 2ab \cdot \cos \widehat{C}.$$

As mesmas figuras nos dão, no primeiro caso:

$$h = c \cdot \operatorname{sen} \widehat{B} = b \cdot \operatorname{sen} \widehat{C},$$

logo

$$\frac{b}{\operatorname{sen} \widehat{B}} = \frac{c}{\operatorname{sen} \widehat{C}}.$$

No segundo caso temos

$$h = b \cdot \operatorname{sen} \widehat{C}$$

e

$$h = c \cdot \operatorname{sen}(\pi - \widehat{B}) = c \cdot \operatorname{sen} \widehat{B},$$

logo, novamente:

$$\frac{b}{\operatorname{sen} \widehat{B}} = \frac{c}{\operatorname{sen} \widehat{C}},$$

como antes.

Se tomarmos a altura baixada do vértice B sobre o lado AC, obteremos, com o mesmo argumento, a relação

$$\frac{a}{\operatorname{sen} \widehat{A}} = \frac{c}{\operatorname{sen} \widehat{C}}.$$

Podemos então concluir que, em qualquer triângulo, tem-se

$$\frac{a}{\operatorname{sen} \widehat{A}} = \frac{b}{\operatorname{sen} \widehat{B}} = \frac{c}{\operatorname{sen} \widehat{C}}.$$

Esta é a *lei dos senos*. Ela diz que, em todo triângulo, a razão entre um lado e o seno do ângulo oposto é constante, isto é, é a mesma seja qual for o lado escolhido. Há uma interpretação geométrica para a razão $a/\operatorname{sen}\widehat{A}$. Ela é igual ao diâmetro do círculo circunscrito ao triângulo ABC.

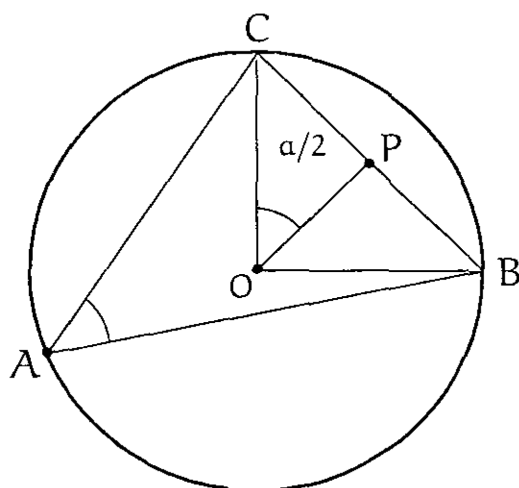


Figura 91

Com efeito, a perpendicular OP, baixada do centro do círculo circunscrito sobre o lado BC é também mediana do triângulo isósceles OBC e bissetriz do ângulo $C\widehat{O}B$, que é igual a $2\widehat{A}$. Logo $C\widehat{O}P = \widehat{A}$ e daí resulta que $\frac{a}{2} = r\operatorname{sen}\widehat{A}$, ou seja, $\frac{a}{\operatorname{sen}\widehat{A}} = 2r =$ diâmetro do círculo circunscrito ao triângulo ABC.

As leis dos cossenos e dos senos permitem obter os seis elementos de um triângulo quando são dados três deles, desde que um seja lado, conforme os casos clássicos de congruência de triângulos.

Problema. Determinar, no triângulo ABC, os lados a, b, c e os ângulos $\widehat{A}$, $\widehat{B}$, $\widehat{C}$ nos seguintes casos:

1. *São dados os lados* a, b, c.

Então

$$a^2 = b^2 + c^2 - 2bc\cos\widehat{A},$$

logo

$$\cos\widehat{A} = \frac{b^2 + c^2 - a^2}{2bc}$$

e isto nos permite determinar $\widehat{A}$.

Analogamente se obtém o ângulo $\widehat{B}$; o ângulo $\widehat{C}$ pode ser mais facilmente obtido a partir da relação $\widehat{A} + \widehat{B} + \widehat{C} = 2$ retos.

Observação. Para que exista um triângulo com lados $a \leqslant b \leqslant c$ é necessário e suficiente que se tenha $c < a + b$.

2. *São dados os lados* a, b *e o ângulo* $\widehat{C}$.

Neste caso, o lado c se obtém pela lei dos cossenos:

$$c = \sqrt{a^2 + b^2 - 2ab\cos\widehat{C}},$$

recaindo-se assim no caso anterior.

3. *São dados os ângulos* $\widehat{A}$, $\widehat{B}$ *e o lado* c.

Determina-se o ângulo $\widehat{C}$ pela igualdade $\widehat{A} + \widehat{B} + \widehat{C} = 2$ retos e o lado a pela lei dos senos, segundo a qual $a/\operatorname{sen}\widehat{A} = c/\operatorname{sen}\widehat{C}$, logo $a = c \cdot \operatorname{sen}\widehat{A}/\operatorname{sen}\widehat{C}$. Agora tem-se os lados a, c e o ângulo $\widehat{B}$ formado por eles. Recai-se assim no caso anterior.

Observação. Para que $\widehat{A}$ e $\widehat{B}$ sejam ângulos de um triângulo, é necessário e suficiente que $\widehat{A} + \widehat{B} < 2$ retos.

4. *São dados os lados* a, b, *com* $a > b$, *e o ângulo* $\widehat{A}$.

Este é o pouco conhecido quarto caso de congruência de triângulos, segundo o qual dois triângulos são congruentes quando têm dois lados iguais e um ângulo igual oposto ao maior desses dois lados. Note-se que $\widehat{A} > \widehat{B}$, logo o ângulo $\widehat{B}$ é agudo.

Aqui se usa novamente a lei dos senos. A partir da proporção

$$\frac{a}{\operatorname{sen}\widehat{A}} = \frac{b}{\operatorname{sen}\widehat{B}} \quad \text{obtém-se} \quad \operatorname{sen}\widehat{B} = \frac{b}{a}\operatorname{sen}\widehat{A}.$$

Como $b < a$, vemos que $\frac{b}{a}\operatorname{sen}\widehat{A}$ é um número positivo menor do que 1, logo existe um único ângulo $\widehat{B}$, menor do que dois retos, cujo seno é igual a $\frac{b}{a}\operatorname{sen}\widehat{A}$. Em seguida, determina-se o ângulo $\widehat{C}$ pela igualdade $\widehat{A}+\widehat{B}+\widehat{C} = 2$ retos. Agora, conhecendo a, b e $\widehat{C}$, recai-se no caso 2.

Observação. Do ponto de vista em que nos colocamos, o triângulo ABC é dado, tratando-se apenas de calcular 3 dos seus elementos quando são dados outros 3. Por isso não cabia acima indagar se $\widehat{A} + \widehat{B} < 2$ retos, antes de calcular $\widehat{C}$. Entretanto, é verdade que, dados $a > b$ e $\widehat{A} < 2$ retos, existe um triângulo ABC tal que $\overline{BC} = a$, $\overline{AC} = b$ e $\widehat{A}$ é o ângulo dado. Para ver isto, tome um segmento AC de comprimento b e uma semi-reta AX tal que o ângulo $C\widehat{A}X$ seja igual ao ângulo $\widehat{A}$ dado. Com centro no ponto C, trace uma circunferência de raio a. Como $b < a$, o ponto A pertence ao interior dessa circunferência, logo a semi-reta AX corta a circunferência num único ponto B, que é o terceiro vértice do triângulo procurado.

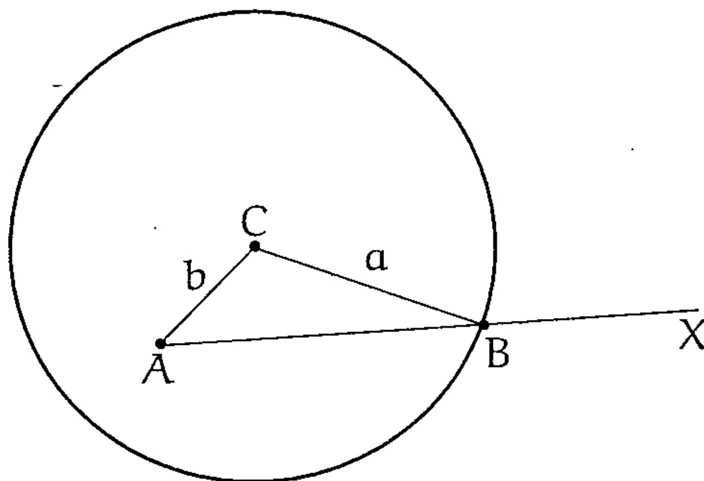


Figura 92

O programa de Matemática da primeira série do Ensino Médio tem como tema central as funções reais de uma variável real, estudadas sob o ponto de vista elementar, isto é, sem o uso do Cálculo Infinitesimal. Como preliminar a esse estudo e preparação para as séries subseqüentes, são apresentadas noções sobre conjuntos, a idéia geral de função e as diferentes categorias de números: naturais, inteiros e, principalmente, reais.

O presente livro cobre esse programa. Ele foi escrito para professores de Matemática do Ensino Médio e para estudantes de licenciatura em Matemática. Sua principal mensagem é de que os conceitos abstratos da Matemática servem de modelos para situações concretas, permitindo analisar, prever e tirar conclusões em circunstâncias onde uma abordagem empírica é insatisfatória.

A fim de saber que espécie de função se deve empregar para resolver um determinado problema, é necessário conhecer as propriedades características de cada função, pois as situações da vida real, quer no cotidiano, quer na Tecnologia, quer na Ciência, não surgem acompanhadas de fórmulas explícitas. Este é um ponto de fundamental importância, freqüentemente ignorado no ensino formal tradicional, onde os conceitos matemáticos são introduzidos para resolver problemas que se referem a eles mesmos.

Neste livro, as funções dos vários tipos apresentados são caracterizadas por meio de propriedades simples, permitindo deste modo ao professor empregá-las conscientemente.

8ed.(200

A ma

ISBN 85-85818-10-7



